EXHIBIT 41
[FILED UNDER SEAL]

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF TEXAS
SHERMAN DIVISION

The State of Texas, et. al.
Plaintiff,

v.

Google LLC,
Defendant.

Case No: 4:20-cv-00957

EXPERT REBUTTAL REPORT OF
CYNTHIA RUDIN

09.09.2024

Cynthia Rudin

Cynthia Rudin

Table of Contents

I. INTRODUCTION

A. Assignment

1. I understand that on December 16, 2020, a multistate coalition led by the State of Texas filed a lawsuit against Google LLC ("Google") asserting violations by Google of federal and state antitrust laws and violations of other state laws, in connection with Google's conduct in the online display advertising industry and as to digital advertising technologies ("Ad Tech" or "Ad Tech stack"). Currently, 16 States (Texas, Alaska, Arkansas, Florida, Idaho, Indiana, Kentucky, Louisiana, Mississippi, Missouri, Montana, Nevada, North Dakota, South Carolina, South Dakota, and Utah) and the Territory of Puerto Rico are Plaintiffs in the case ("Plaintiff States"). I have been retained to provide expert analysis and opinions on behalf of all of the Plaintiff States. [1]

2. I have been asked by counsel for the State of Texas, on behalf of all Plaintiff States in this case, to evaluate certain opinions discussed in the expert reports of Prof. Paul R. Milgrom,

[1] Engagement Letter signed on July 05, 2021.

Prof. Anindya Ghose, Prof. Steven N. Wiggins, and Prof. Michael R. Baye that were submitted on 2024.07.30 and 2024.08.06.[2]

3. I am being compensated at the rate of $550 per hour for my work on this matter. I have been supported by a research team who has provided research support and assistance in my preparation of this report under my supervision, direction, and instruction. My compensation does not depend on the opinions or testimony that I may give or on the outcome of this case.
4. I have signed the Protective Order on 7/15/2021. The list of documents and materials I have considered and relied upon are cited inline throughout this Rebuttal Report. The list is also provided in Appendix B.

5. A list of all documents, including transcripts, considered in this report is attached in Appendix B. I understand that document productions are ongoing in this case and that additional relevant documents may be produced in this case by Google and third parties right before and after I issue this report. I also understand that, after I submit this Rebuttal Report, expert and fact witnesses for Google and the U.S. Department of Justice and other plaintiffs will be testifying at trial in the parallel case pending in the Eastern District of Virginia (*United States et al. v. Google LLC, No. 1:23-00108*). I may, and reserve the right to, review and rely on additional documents, including transcripts, and testimony in conducting my work and forming my opinions in this case. I reserve the right to supplement or amend this report if my opinions change or require supplementation as a result of my ongoing review of documents and to create and use graphics, figures, and/or other illustrations at trial to support my conclusions

B. Qualifications

---

[2] *See* Expert Report of P. Milgrom, ¶32 (“[A]dvertisers’ and publishers’ routine data analysis and experimentation with bids and floor prices are typically sufficient for them to identify optimal strategies.”)

See Expert Report of S. Wiggins, ¶¶ 40-41 ("Plaintiffs’ theory further fails to take into account the evidence that advertisers and publishers learn through experimentation and rely on numerous specialized intermediaries to help them implement optimal strategies... In Section II.B.1, I discuss the economic literature on learning-by-doing. Section II.B.2 describes various features of the Ad Tech industry that facilitate advertiser learning, as well as the evidence regarding how advertisers generally learn and seek to maximize their returns. In Section II.B.3, I discuss the evidence of similar behavior by publishers.”)

See Expert Report of A. Ghose, ¶134 ("[I]t is well understood in the industry that ad tech providers conduct experiments and optimizations, and indeed, publishers and advertisers expect their ad tech providers to do so.")

See Expert Report of Michael R. Baye, Exhibit 11 “Impressions on DFP for Display Ads (Narrow) Plus Direct Deals Viewed by U.S. Users Excluding House Ads June 2013 – March 2023.”

6. My name is Cynthia Rudin. I am the Gilbert, Louis, and Edward Lehrman Distinguished Professor of Computer Science at Duke University, and I have appointments in the Departments of Computer Science, Electrical and Computer Engineering, Statistical Science, Mathematics, and Biostatistics & Bioinformatics at Duke.
7. My core expertise is in machine learning, statistics, optimization, and data science. The goal of statistics is to quantify uncertainty. Machine learning is a type of predictive statistics. I deal with problems of estimation and uncertainty quantification every day.
8. I am the sole recipient of the 2022 Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity from the Association for the Advancement of Artificial Intelligence (AAAI). This award is the most prestigious award in the field of artificial intelligence. Similar only to world-renowned recognitions such as the Nobel Prize and the Turing Award, it carried a monetary reward at the million-dollar level.
9. I am also the only three-time winner of the INFORMS Innovative Applications in Analytics Award. The first time I won this, it was for work I did on data analytics for power grid reliability in New York City; I led the first major effort to maintain an underground electrical distribution network using machine learning. This was joint work with Con Edison in New York City. The second time I won this award was for work on early detection of cognitive decline (Alzheimer's, dementia, and Parkinson's) by analyzing drawings that patients perform using a digital pen. The third time I won this award, it was for developing a medical scoring system called the 2HELPS2B score that is widely used in hospitals to assess seizure risk in intensive care unit patients.
10. My collaborators and I also developed code for detecting crime series in cities. This methodology (specifically, the Series Finder algorithm) was developed in Cambridge, MA and adapted by the NYPD. The NYPD's application of the methodology (Patternizr) has been running live in New York City since 2016 for determining whether each new crime is related to past crimes.
11. I achieved second place in the 2023 Bell Labs Prize competition, earning a $50,000 prize for work that "recognizes game-changing innovations in science, technology, engineering and mathematics."[3] Our entry focused on algorithms for interpretable machine learning that produce predictive formulas that are simple enough to print on an index card.

[3] Nokia Bell Labs. Bell Labs Prize, Accessed on September 6, 2024. https://www.bell-labs.com/collaboration-opportunities/prize/

12. My lab's PaCMAP algorithm for dimension reduction for data visualization received the 2023 John M. Chambers Statistical Software Award, as well as the 2024 Award for Innovation in Statistical Programming and Analytics (AISPAA) from the American Statistical Association.
13. I received the Outstanding Paper Award from the Association for Computational Linguistics in 2023 for work on computationally generated poetry. I am a winner of the 2021 "Best OM in OR Award" from INFORMS in 2021, which is given to the best operations management paper in Operations Research. Numerous additional best paper awards can be found in my CV or on my website.
14. I am heavily involved in five scientific communities, each having separate professional meetings: machine learning, artificial intelligence, data mining, management science (specifically, the Institute for Operations Research and Management Science - INFORMS), and the American Statistical Association. I am a past chair of both the INFORMS Data Mining Section and the Statistical Learning and Data Science Section of the American Statistical Association. I have served on committees for DARPA, the National Institute of Justice, the Association for the Advancement of Artificial Intelligence (AAAI), and ACM SIGKDD (Association for Computational Machinery, Special Interest Group on Knowledge Discovery in Data). I am presently on the Executive Committee for ACM SIGKDD.
15. I have served on four committees for the National Academies of Sciences, Engineering and Medicine, including the Committee on Applied and Theoretical Statistics, the Committee on Law and Justice, the Committee on Analytic Research Foundations for the Next-Generation Electric Grid, and the National Academies Committee on Facial Recognition Technology.
16. I have given keynote/invited talks at several conferences, including the top conferences in data mining (KDD, ECML-PKDD, ICDM), top conferences in machine learning (AISTATS – Artificial Intelligence and Statistics), INFORMS, Machine Learning in Healthcare (MLHC), Fairness, Accountability and Transparency in Machine Learning (FAT-ML), the SPIE Medical Imaging Conference, and the Nobel Conference.
17. I was named as one of the "Top 40 Under 40" by Poets and Quants in 2015 and was named by Businessinsider.com as one of the 12 most impressive professors at MIT in 2015.
18. I am a 2022 Guggenheim fellow, as well as a fellow of the American Statistical Association, the Institute of Mathematical Statistics, and AAAI.

19. I enjoy coaching teams to enter data science competitions, and we have won many of them, including the FICO Recognition Award for the Explainable Machine Learning Challenge (2018) on the topic of loan default prediction, NTIRE Superresolution Competition (2018) on the topic of image superresolution, PoeTix Literary Turing Competition (2018) on automatically generated poetry, and American Statistical Association Data Challenge Expo Student Competition (2022 and 2023).
20. I publish extensively in the top venues in machine learning, artificial intelligence, data mining, and computer vision. I have published also in the top journals in operations research, applied statistics, medical informatics, music information retrieval, several areas of medicine (including an article in Radiology in 2024, as well as in neurology, sleep medicine, infectious disease, etc.), and the Proceedings of the National Academy of Sciences. According to Google Scholar, I have over 24,800 citations, an h-index of 56, which means 56 of my papers have been cited over 56 times, and an i-10 index of 134, which means 134 of my papers have been cited over 10 times.
21. I have written a publicly available textbook on machine learning that is self-published on my website. It is called "Intuition for the Algorithms of Machine Learning," and it serves as the graduate textbook for the class I teach every fall to approximately 200 students at Duke University.

C. Summary of opinions

22. I have reviewed the expert reports of Prof. Milgrom, Prof. Ghose, Prof. Wiggins, and Prof. Baye and applied my training and experience to analyze Google's machine learning programs from a statistical and optimization lens to offer the following opinions in my report. My opinions apply to various Google programs, including Google's buy-side Dynamic Revenue Sharing ("DRS"), DRS in AdX, Project Bernanke, Enhanced Dynamic Allocation ("EDA"), Reserve Price Optimization ("RPO"), a shift from second-price auctions to first-price auctions, Open Bidding, Unified Pricing Rules ("UPR"), and Alchemist, which I collectively refer to as "Google's auction manipulations."

23. I respond to some of Google's experts' opinions, including:

a. Prof. Milgrom's opinion that "advertisers' and publishers' routine data analysis and experimentation with bids and floor prices are typically sufficient for them to identify optimal strategies."[4]

b. Prof. Wiggins' opinion that "advertisers and publishers can learn through experimentation."[5]

c. Prof. Ghose's opinion that "[w]ith digital advertising, advertisers have access to large amounts of data, sophisticated analytics tools, and real-time insights, which enable them to make more informed decisions about budget allocation."[6]

d. Prof. Ghose's opinion that Google's auction manipulations were disclosed appropriately.[7]

24. Prof. Ghose's opinion that "data have diminishing returns to scale—in other words, incremental additions of data are increasingly less valuable—which suggests that companies with large amounts of user data are not necessarily more competitive than companies with relatively less user data."[8]

25. Contrary to Google's experts' opinions, due to Google's auction manipulations, sellers and buyers would not be able to experiment to arrive at optimal strategies for at least the following reasons:

a. Google's auction manipulations would have interfered with publishers' and advertisers' natural optimization goals.

---

[4] Expert Report of P. Milgrom, ¶32. Here, optimal strategies refer to the "process of setting reserve prices" for publishers and "determining bids" for advertisers.

[5] Expert Report of S. Wiggins, ¶¶ 40-41 ("Plaintiffs' theory further fails to take into account the evidence that advertisers and publishers learn through experimentation and rely on numerous specialized intermediaries to help them implement optimal strategies... In Section II.B.1, I discuss the economic literature on learning-by-doing. Section II.B.2 describes various features of the Ad Tech industry that facilitate advertiser learning, as well as the evidence regarding how advertisers generally learn and seek to maximize their returns. In Section II.B.3, I discuss the evidence of similar behavior by publishers.")

[6] Expert Report of A. Ghose, ¶105 ("With digital advertising, advertisers have access to large amounts of data, sophisticated analytics tools, and real-time insights, which enable them to make more informed decisions about budget allocation."); *id.* at ¶109 ("These data provide specific and actionable insights, allowing advertisers to adapt quickly and optimize their campaigns. Traditional analytical methods, as well as machine learning and A/B testing, are increasingly used to analyze digital advertising data. Digital advertising also enables more precise targeting, ensuring that ads reach specific, defined audiences.")

[7] Expert Report of A. Ghose, ¶145 ("Plaintiff's experts fail to acknowledge that Google did indeed disclose its experiments and auction rules, and that any more detailed disclosures by Google or other ad tech providers would run a number of risks.")

[8] Expert Report of A. Ghose, ¶185 ("Plaintiffs' experts ignore academic research demonstrating that data have diminishing returns to scale—in other words, incremental additions of data are increasingly less valuable—which suggests that companies with large amounts of user data are not necessarily more competitive than companies with relatively less user data.")

e. Google's auction manipulations (which were oft changing) and their unknown degrees of freedom would have prevented publishers and advertisers from experimenting effectively to detect and mitigate those programs.

b. Google actively concealed Google's auction manipulations and failed to disclose their mechanics, and, in any event, its generic disclosures did not mitigate the data sufficiency problems buyers and sellers would have faced.

c. Contrary to Prof. Ghose's opinion, publishers and advertisers have access to substantially less data for machine learning than Google does, including the sufficient "fresh" data required to make an effective machine learning model of auction behavior.

d. Google's auction manipulations further limit the amount of useful, fresh information available to publishers and advertisers.

26. My analysis shows that access to less useful data, the unknown degrees of freedom introduced by Google's auction manipulations and experiments, Google's insufficient disclosures of its auction manipulations, and Google's repeated efforts to thwart publishers' and advertisers' efforts to gain more information would have prevented publishers and advertisers from experimenting effectively to find optimal strategies or detect and respond to Google's auction manipulations.

27. My analysis demonstrates that publishers and advertisers also would not have been able to successfully respond to Google's auction manipulations, even if they were to deploy machine learning models of their own, because Google's auction manipulations would have created a compounding or snowball effect that would enable Google products to have a higher chance to win more auctions while also improving Google's internal machine learning models. This feedback loop would have given Google access to more data, allowing the company to positively reinforce outcomes to its benefit.

II. MACHINE LEARNING BACKGROUND AND FEASIBILITY OF MACHINE LEARNING TO INFORM AUCTION STRATEGY

28. Machine learning (also referred to as "ML"), as the name suggests, is a branch of artificial intelligence that enables machines to learn from data to uncover hidden patterns and make

predictions about the patterns that will emerge in new or similar data.[9] ML is applied in various use cases such as image and speech recognition, recommendation systems, cybersecurity and threat detection, user behavior analytics, and natural language processing, among others.[10]

29. ML is ideally suited for situations where the volume of data required to detect patterns is too large to identify them using other tools. Analyzing the behavior of participants in an auction is an example of an ideal ML use case, because each auction may involve different parties, different budgets, different products, and different levels of auction experience. Identifying trends in auctions as they are affected by these and other variables is too complex to analyze with paper and pencil, or even with standard spreadsheet operations.

30. I have been asked in this case to review the feasibility of buyers or sellers using ML to guide their auction strategies when using Google's online advertising tools. In my opinion, it would not be feasible for buyers or sellers to use ML to effectively guide their auction strategies in the context of Google's auction manipulations as explained in Sections III and V. In this section, I provide a summary of key ML concepts relevant to my opinions, an overview of how Google uses ML to optimize its display advertising, and a discussion of a hypothetical auction where ML is used to inform buyer or seller strategy.

A. Machine learning concepts relevant to optimizing auction strategy

31. The following terms and concepts are relevant to my opinions responding to Google's experts in this matter.

[9] Nicholas, J., Herbert Chan, H.W., Baker, M.A.B. "Machine learning: applications of artificial intelligence to imaging and diagnosis" Biophysical reviews, 11(1), 111–118. ("Machine learning (ML) is an umbrella term that refers to a broad range of algorithms that perform intelligent predictions based on a dataset. These datasets are often large, perhaps consisting of millions of unique data points. Recent progress in machine learning has attained what appears to be a human level of semantic understanding and information extraction, and sometimes the ability to detect abstract patterns with greater accuracy than human experts.")

[10] Sarker, I.H. "Machine Learning: Algorithms, Real-World Applications and Research Directions" SN Computer Science (2021) 2:160. ("Thus, this study's key contribution is explaining the principles of different machine learning techniques and their applicability in various real-world application domains, such as cybersecurity systems, smart cities, healthcare, e-commerce, agriculture, and many more. … Image, speech and pattern recognition: Image recognition is a well-known and widespread example of machine learning in the real world, which can identify an object as a digital image. For instance, to label an x-ray as cancerous or not, character recognition, or face detection in an image, tagging suggestions on social media, e.g., Facebook, are common examples of image recognition. Speech recognition is also very popular that typically uses sound and linguistic models.")

32. **Machine Learning Models:** ML models are computer programs that are used to recognize patterns in data and/or make predictions based on previously analyzed data.[11] For this matter, I focus on *supervised* ML, where models learn to use inputs ("features" x) to predict outcomes (for instance, outcome y) based on training data consisting of (x,y) pairs.[12]
33. **Training Data:** To make accurate predictions, ML models must be given sufficient training data to learn relationships between features x and outcomes y. The models learn patterns from examples in these training datasets to generalize and make forecasts about future outcomes.[13]
34. **Fresh Data:** Good quality training data for ML models includes relevant data at a sufficient scale, or volume, that is reasonably up to date (or "fresh" data).[14] Fresh data can provide a more accurate understanding of a given system. As an example, in order to estimate the gas prices in an area tomorrow, it would be more helpful to have data about gas prices yesterday than data about what they were one year ago.
35. **Estimations:** In many industries, people need to estimate or make predictions based on data. Estimations use data from a sample to approximate a value for a larger population.[15]
36. **Variance**: Estimation becomes more difficult when noise (i.e. uncertainty) is present in the data. Data scientists refer to this as "variance." Many basic inequalities in probability[16] are

---

[11] Lecture notes of Rudin, C. "Fundamentals of Learning Course Notes" from Intuition for the Algorithms of Machine Learning: A Multimedia Textbook. Duke Computer Science, https://users.cs.duke.edu/~cynthia/CourseNotes/ConceptsOfLearningNotes.pdf ("Machine learning is a broad field within machine learning and predictive statistics. It involves the design of algorithms that learn by example. Machine learning is (in a broad sense) pattern recognition.")

[12] Cunningham, P., Cord, M., Delany, S.J "Supervised Learning " Machine Learning Techniques for Multimedia. Cognitive Technologies. 11. ("The defining characteristic of supervised learning is the availability of annotated training data. The name invokes the idea of a 'supervisor' that instructs the learning system on the labels to associate with training examples. Typically these labels are class labels in classification problems. Supervised learning algorithms induce models from these training data and these models can be used to classify other unlabeled data.")

[13] Sarkis, A. "Training Data for Machine Learning: Human Supervision from Annotation to Data Science" O'Reilly Media, Inc. ("Code explicitly defines logic and operations, whereas training data provides examples that the model must generalize from.")

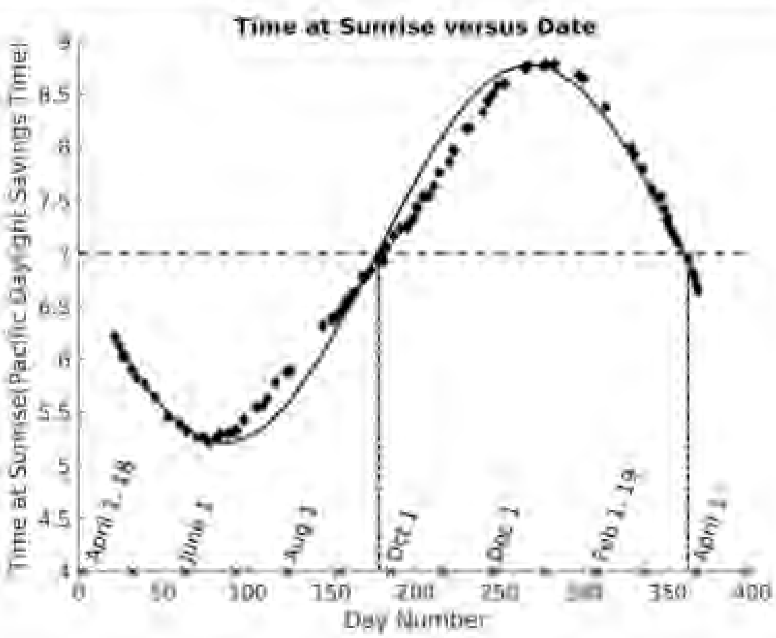
[14] Gudivada, V., Apon, A., and Ding, J. "Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations" International Journal on Advances in Software, vol 10 no 1 & 2, year 2017. 1. ("What is data quality? It depends on the task and is often defined as the degree of data fitness for a given purpose. It indicates the degree to which the data is complete, consistent, free from duplication, accurate and timely for a given purpose.")

[15] Singh, Ajay S., and Micah B. Masuku. "Sampling techniques & determination of sample size in applied statistics research: An overview." *International Journal of economics, commerce and management* 2.11 (2014): 1-22. ("Sampling is related with the selection of a subset of individuals from within a population to estimate the characteristics of whole population.… Generally different sampling methodologies help to draw the good sample or better representative for estimation of parameters. Sample size is also more important to increase the precision of results, minimize the variability and for generalization of results with interpretation.")

[16] ScienceDirect. "Chebyshev Inequality", Accessed on September 1, 2024. https://www.sciencedirect.com/topics/mathematics/chebyshev-inequality;

fundamentally based on the fact that higher variance signals lead to harder estimation problems that degrade the performance of learning algorithms.[17] When variance is low, forecasting can be easier—e.g., forecasting the minute of the sunrise, at some fixed location, based on time of year, *see* Figure 1. On the other hand, high-variance time series—such as stock prices, outdoor temperature measurements, or electroencephalogram (EEG) measurements—tend to be hard to predict, *see* Figure 2.[18]

Figure 1: Estimating daily time at sunrise (low variance)[19]



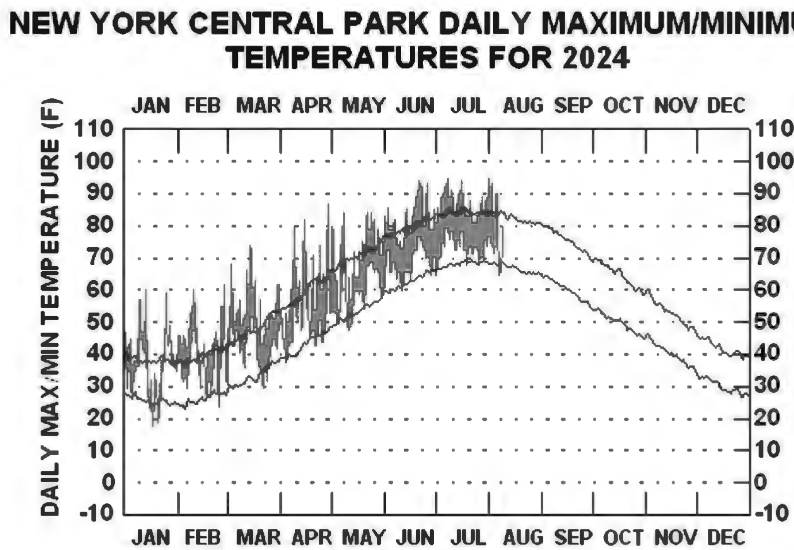
Lecture notes of Paninski, L. "Probability inequalities " (Columbia University Course) Statistics 4109: Probability and Statistics (Fall 2010). http://www.stat.columbia.edu/~liam/teaching/4109-fall10/inequalities-notes.pdf ("[A] probability inequality [is] a bound on the probability of some undesired event happening.")

[17] Gupta, S., and Gupta, A. "Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review" Procedia Computer Science 161 (2019) 466–474. ("The occurrences of noisy data in data set can significantly impact prediction of any meaningful information. Many empirical studies have shown that noise in data set dramatically led to decreased classification accuracy and poor prediction results. Therefore, the problem of identifying and handling noise in prediction application has drawn considerable attention over past many years.")

[18] Golestani, A., Gras, R. "Can we predict the unpredictable?" Scientific Reports 4, Article Number: 6834 (2014). 1. ("Time series forecasting is of fundamental importance for a variety of domains including the prediction of earthquakes, financial market prediction, and the prediction of epileptic seizures. We present an original approach that brings a novel perspective to the field of long-term time series forecasting. Nonlinear properties of a time series are evaluated and used for long-term predictions. We used financial time series, medical time series and climate time series to evaluate our method. The results we obtained show that the long-term prediction of complex nonlinear time series is no longer unrealistic.")

[19] Figure from Sky & Telescope The Essential Guide to Astronomy. "95 Sunrises Along the Columbia Riverbank" (May 28, 2020), Accessed on August 30, 2024. https://skyandtelescope.org/observing/stargazers-corner/95-sunrises/

Figure 2: Estimating daily temperatures (high variance)[20]



37. **Number of Variables Considered:** In statistics and ML, the number of variables, also called the number of degrees of freedom, plays a critical role in estimation, prediction, and optimization.[21] The number of degrees of freedom is a major consideration for every statistical model because of the "Curse of Dimensionality," which states that the amount of data needed for modeling increases exponentially with the number of degrees of freedom.[22]

38. Although ML models can learn based on reasonably sized real-world datasets, even without an exponential amount of data per variable, they struggle with each additional possible variable included in the model. Even if the individual constructing a ML model restricts the predictive

---

[20] Figure from ClimateStations.com. "ny2024.gif", Accessed on August 30, 2024. https://www.climatestations.com/wp-content/uploads/2024/08/ny2024.gif

[21] Yu, C.H. "Degrees of Freedom" International Encyclopedia of Statistical Science. 363. ("Many elementary statistics textbooks introduce the concept of degrees of freedom (df) in terms of the number scores that are 'free to vary.' However, this explanation cannot clearly show the purpose of df. There are many other approaches to present the concept of degrees of freedom. Two of the most meaningful ways are to illustrate df in terms of sample size and dimensionality. Both represent the number of pieces of useful information.")

[22] Lecture notes of Strohmer, T., "Chapter 2: Curses, Blessings, and Surprises in High Dimensions" UC Davis, Mathematics https://www.math.ucdavis.edu/~strohmer/courses/270/Chapter2.pdf ("The curse of dimensionality refers to the fact that many algorithmic approaches to problems in Rd become exponentially more difficult as the dimension d grows." (Here, Rd means a d-dimensional space, i.e., d degrees of freedom.))

Berisha, V., Krantsevich, C., Hahn, P.R., Hahn, S., Dasarathy, G., Turaga, P., Liss, J. "Digital medicine and the curse of dimensionality" npj Digital Medicine volume 4, Article number: 153 (2021). 1. ("As we increase the number of clinical variables we measure, there is a combinatorial explosion in the possible values that the variables can jointly take. Building robust models for solving complex problems requires that the increase in variability is offset by a commensurate increase in sample size. Attempting to solve highly complex real-world problems using high-dimensional data, without increasing sample size, leaves datasets with a 'blind spot' - contiguous regions of feature space without any observations - and poses several challenges to model development. This phenomenon is known as the curse of dimensionality in statistical learning theory.")

models to be simpler and less computationally complex, it is the number of variables (the number of degrees of freedom) being considered that makes getting accurate predictions/forecasts challenging for the modeler. The Curse of Dimensionality affects ML modeling, since the number of variables in the model is the number of degrees of freedom – each additional variable added to the model increases the complexity of the model, requiring an increased number of data points to train the model.[23]

39. **A/B testing** – A/B testing is a way to compare two versions (A and B) of something to understand which version performs better.[24] ML model developers can employ A/B testing to decide which version of the model performs better in real- world scenarios. They would run experiments trying both versions A and B on a sufficiently large sample to see which leads to superior outcomes according to an objective.[25]

B. The accuracy of machine learning models is primarily driven by the usefulness of their training data

40. In addition to understanding certain terminology, it is important to understand that any particular machine learning use case requires sufficient usable data. In the context of using machine learning for auction strategy optimization that Google's experts suggest, such machine learning algorithms would need enough relevant data to uncover underlying patterns that occur in the auction context.

41. The performance of a machine learning model depends, to a great degree, on the quality of training data it receives. High-quality data with powerful variables leads to more accurate and precise models, whereas a training dataset that does not represent the problem at hand will produce poor predictions. In this report, I use "data freshness" to indicate how recently the data

---

[23] Altman, N. and Krzywinski, M. "The curse(s) of dimensionality" Nature Methods. ("However, in the 'big data' era, the sheer number of variables that can be collected from a single sample can be problematic. This embarrassment of riches is called the 'curse of dimensionality' (CoD) and manifests itself in a variety of ways. This month, we discuss four important problems of dimensionality as it applies to data sparsity, multicollinearity, multiple testing and overfitting. These effects are amplified by poor data quality, which may increase with the number of variables.")

[24] Kohavi, R., Tang, D., Xu, Y. "Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing " Cambridge University Press 2020. 7. ("In a simple A/B test, A and B are the two variants, usually called Control and Treatment. In some literature, a variant only means a Treatment; we consider the Control to be a special variant: the existing version on which to run the comparison.")

[25] Censius Blog. "How To Conduct A/B Testing In Machine Learning?", Accessed on September 4, 2024. https://censius.ai/blogs/how-to-conduct-a-b-testing-in-machine-learning ("Using the A/B testing approach, machine learning models may be evaluated and improved. The approach may be used to see if a new model is better than one already exists. The organization should choose a metric to compare the control and new models for this purpose. This metric is used to assess deployment success and differentiate between the two.")

were collected. ML models are "data hungry"[26] in that they benefit from larger datasets with more variables.

42. The relevancy, size, granularity, and freshness of data for the different machine learning use cases are discussed below. To illustrate this, I compare the kind of data needed to predict the weather to the kind of data relevant for auctions.

43. Data *freshness* is important, in that more recent information is more valuable in predicting the weather than older data. Knowing the weather one hour ago is more helpful in predicting what the weather will be in one hour than having the knowledge of the weather yesterday, last week, or last year. Herein, I make the reasonable assumption that the same freshness of data holds value in display ad auctions as well, i.e., knowing that a person has recently visited a website for electronics like Samsung or Sony within the last hour is more valuable to decide whether to show an ad to them than knowing that this occurred one day or one year ago.

44. Data *relevancy* is important. Again, using weather as an example, you can predict rainfall within the next hour using rainfall in your exact location from 10 minutes ago, but a more powerful predictor would use all the rainfall sensors in the path of a storm coming toward you at 10-minute intervals over the last half hour. That set of variables for prediction is more powerful because you can see whether rainfall is heading your way, even if it is not raining at your location now. Herein, I make the reasonable assumption that the same is true for display ad auctions as well, i.e., a seller trying to optimize floor prices (for example, to raise revenue, increase numbers of impressions sold, or improve the quality of ads shown) could try to predict what each buyer might bid using just data from their own past auctions. However, a substantially more comprehensive view of each buyer would be obtained by having access to their bids for all auctions over all publishers. Similarly, a buyer might want to optimize their bid to outbid others but might only have bid in a similar auction a few times recently. Having winning bids from all similar recent auctions would be valuable in optimizing the buyer's new bid.

---

[26] van der Ploeg, T., Austin, P.C. & Steyerberg, E.W. "Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints" BMC Medical Research Methodology volume 14, Article number: 137 (2014). ("Modern modelling techniques such as SVM, NN and RF need far more events per variable to achieve a stable AUC-value than classical modelling techniques such as LR and CART. If very large data sets are available, modern techniques such as RF may potentially achieve an AUC-value that exceeds the AUC-values of modelling techniques such as LR. The improvement over simple LR models may, however, be minor, as was shown in the two empirical examples in this study. This implies that modern modelling techniques should only be considered in medical prediction problems if very large data sets with many events are available.")

45. Data *size* is important. If the individual constructing a ML model wanted to predict rainfall next month, using only one hour's worth of data from last month is not as valuable as using data from every hour last month. Herein, I make the reasonable assumption that the same is true for display ad auctions as well, i.e., the more past auction data is available, the easier it is for buyers to determine the value of the ad (their ROI) and the amount they would want to bid in order to outbid other bidders, and the easier it is for sellers to set price floors.

46. Prof. Ghose's remark that "data have diminishing returns to scale—in other words, incremental additions of data are increasingly less valuable—which suggests that companies with large amounts of user data are not necessarily more competitive than companies with relatively less user data" does not apply to machine learning in general nor to buyers'/sellers' need for data in auctions.[27] Prof. Ghose's opinion that data scale has diminishing returns is like saying that you get diminishing returns for predicting weather within the next hour when you only receive more and more weather data from before last year. Yet it is today's data that is important for predicting weather. In fact, the last hour's data will help you predict better, and the last 10 minutes will help even more. In other words, Prof. Ghose has confused the amount of data with the freshness and relevancy of it. In one internal presentation, a Google employee indicated that Google highly valued fresh data in its algorithms, stating that "[f]resh data and fresh stats means that you need to rely less of trying to predict the distant future and instead can see how things are actually performing right now based on concrete data… [y]ou wouldn't drive a car based on traffic on the road from yesterday, you want to know what happened now. [redacted][28]

[27] Expert Report of A. Ghose, ¶185 ("Plaintiffs' experts ignore academic research demonstrating that data have diminishing returns to scale—in other words, incremental additions of data are increasingly less valuable—which suggests that companies with large amounts of user data are not necessarily more competitive than companies with relatively less user data.")

[28] GOOG-NE-02215711 at '717 ("Fresh data and fresh stats means that you need to rely less of trying to predict the distant future and instead can see how things are actually performing right now based on concrete data.")

Ibid. at '718, No Title (No Date) - Internal Google notes ("As I mentioned, one of our core philosophies is the value of fast, fresh data. Faster data means better awareness of what actually happened right now, not what happened yesterday. You wouldn't drive a car based on traffic on the road from yesterday, you want to know what happened now. [redacted]")

47. For instance, one of Google's auction manipulations programs, RPO, required data to be as fresh as [redacted].[29] Old data would not be useful for a model like that, and one internal document stated, "requests for the same cookie are highly localized in time making it hard to train a model based on historical data."[30] As further evidence, Google's machine learning model for Bernanke also needed fresh data.[31]

48. With an increasing number of degrees of freedom comes a need for greater data scale. If the auction is constantly modified by systematic adjustments of price floors, bids and/or take rate, or if throttling is used to conceal how the auction is being run, orders of magnitude more of data would be required for participants to detect the adjustments and respond (assuming it is even possible to respond effectively, which it may not be). For example, in the instance of Google auction manipulations, discussed below, Google's revision of auction floor prices and bids without notice to buyers and sellers makes it practically impossible, or impractically expensive, for a buyer or seller using machine learning to optimize their auction strategies when Google is overriding that strategy in unclear and unpredictable ways.

49. Data advantages can produce a cycle of improvement that accumulates over time, since better data leads to better predictions, which again can lead to collection of more useful data.[32] In the context of applying machine learning to auctions, knowing information about one bid is far less valuable than having information about all bids. Having all bids in all auctions from one seller is much less valuable than having bids from all data from all sellers. This data can be

---

[29] Google had observed that most impression opportunities for a given cookie arrived [redacted]. GOOG-DOJ-14421383 at '383, "UPCOMING LAUNCH - Please review: [Launch 225406] Online Reserve Price Optimization" (February 5, 2018) - Internal email from [redacted] to [redacted]

[30] GOOG-DOJ-14421383 at '383, "UPCOMING LAUNCH - Please review: [Launch 225406] Online Reserve Price Optimization" (February 5, 2018) - Internal email from [redacted] to [redacted] ("The key motivation of online RPO is to use buyers' bidding behavior on high value re-marketing cookies to set reserve prices. While analyzing bidding patterns on cookies, we observed that [redacted] In other words requests for the same cookie are highly localized in time making it hard to train a model based on historical data.")

[31] GOOG-DOJ-15631978 at '979, "Re: Per-buyer dynamic reserve price optimization on AdX - full launch" (November 21, 2014) - Email from [redacted] ("Bernanke uses [redacted]")

[32] Hagiu, Andrei, and Julian Wright. "Data-enabled learning, network effects, and competitive advantage." *The RAND Journal of Economics* 54.4 (2023): 638-667. ("In recent years, much attention has been focused on the role data can play in providing incumbent firms with a competitive advantage. Digitization, connectivity to cloud-based infrastructures, together with cheaper storage and more effective use of data (i.e. improvements in machine learning algorithms), have made it possible for firms in many industries to translate learning from their customer data into rapid improvements in their products. And with better products, these firms can attract more customers (or more usage from existing customers), and therefore obtain more data, potentially creating a self-reinforcing cycle that can make it difficult for any new entrant to compete.")

used to make better predictions of others' bids and estimates of valuations, which lead to more auction wins, which then allows them to collect more data. The model improvements from the data advantage remain with the model, even if the reason for the initial data advantage is later removed.

III. MACHINE LEARNING MODELS IN A HYPOTHETICAL AUCTION BETWEEN BUYERS AND SELLERS

50. In order to opine whether actual publishers and advertisers would have been able to effectively experiment in the context of Google's auction manipulations, below I build a ML framework ("ML Abstracted Auction") to study the problem.
51. I am going to start building my ML Abstracted Auction on a set of postulates or givens. Later below, I will alter or relax some of these givens. I will then connect this abstracted auction to real auctions and opine on whether it would be possible to construct effective ML models in the context of Google's auction manipulations.
52. For each auction, both the buyers and sellers try to estimate the value of the item being auctioned. In my ML Abstracted Auction, it is given that both buyers and sellers will use ML models to get a more precise estimated value of the auction item. They could try to look at similar items and guess what they sold for in the past to estimate value. But they can use sophisticated machine learning methods to automate this process and get better estimates.

A. ML Abstracted Auction with no Auctioneer Manipulations

53. In this hypothetical auction, I assume that the auctioneer's role is to clear the auction and take a percentage of the clearing price. Here, the auctioneer does not use any machine learning to manipulate the auction.

1) Buyer constructs an ML model in my ML Abstracted Auction

54. In my ML Abstracted Auction, the buyer wants to estimate the true value of the auction item.[33] They are unwilling to bid above their valuation and would like to bid as low as possible.

[33] Easley, D. and Kleinber, J. "Networks, Crowds, and Markets: Reasoning about a Highly Connected World" Cambridge University Press, 2010. "Chapter 9: Auctions" p.249 ("The underlying assumption we make when modeling auctions is that each bidder has an intrinsic value for the item being auctioned; she is willing to purchase the item for a price up to this value, but not for any higher price. We will also refer to this intrinsic value as the bidder's true value for the item.")

55. To estimate the value of the auction item in my ML Abstracted Auction, the buyer constructs a ML model that relies on an auction dataset to obtain an estimate. The variables included in the dataset provide detailed information about each auction. I will use an example of an art auction to show some variables a buyer may consider.[34]

a. Who made the auction item ("Who is the artist") - A Latin American artist might be more valuable to an art dealer with Latin American clients, because the dealer can sell to those clients for more;
b. Whether it is the last auction item in a series that a particular buyer owns ("Last piece in Series A by Frida Kahlo") - The buyer may value completing a set more than the buyer values the auction item on its own;
c. Type of auction item ("Painting, sculpture, sketch") - If the buyer is trying to build a sculpture garden, a painting would not be very valuable to them;
d. Dimensions of the auction item ("Size of the painting") - The buyer may be looking to fill a room and needs an exact size;
e. Comparisons ("Price of similar items from a similar time period") - The buyer may consider how similar items have sold before; and
f. Rarity of the auction item ("One of five paintings by an artist") - The buyer may not get another chance to buy the auction item.

56. The buyer uses a machine learning algorithm, which takes the auction dataset and produces a machine learning model for the value of an item. The machine learning model is a formula that combines the variables to estimate the value of the item. It should *generalize*, which means it should accurately predict values for new items, not just the ones in the dataset. The buyer might create a complex transformation of the variables to produce the item's estimated valuation. For instance, it could multiply, add, or exponentiate the variables, compute logical expressions (e.g., "if-then" statements such as "if variable A is 1 and variable B is 2, then produce value 1 and 0 otherwise") and combine them in a nonlinear way to get an estimated value for the item. For this formula to be effective, the buyer needs these variables from past auctions and how much the respective items are sold for.

[34] Note: This exemplary list of variables is illustrative, not exhaustive, of variables that may go into a model's estimation.

57. If the buyer does not have any data, they cannot estimate valuations.
58. The buyer then uses the estimated value of the auction item to create their bid using a bidding model that uses additional variables. Variables that the model would use for bids include:
59. Output from their valuation model;

    a. Allocation and payment rules of the auction, which determine who receives the auction item at what price;[35]
    b. The format of the auction, which can affect the strategy to bid the estimate or bid below the estimate (e.g. 1P versus 2P auctions); [36] and
    c. Other auction format rules, including whether bids are sealed or unsealed, single bid or more complex, and so on. [37]

60. The buyer uses the bidding model to place their bids.

2) Seller constructs an ML model in my ML Abstracted Auction

61. In my ML Abstracted Auction, the seller wants to estimate the value of the auction item as well. They need to estimate the value for each buyer. For instance, one buyer may pay more to complete a collection with the "Last piece in Series A by Frida Kahlo," which would increase the value of the collection as a whole.
62. The seller also needs to estimate the minimum they are willing to accept for the auction item. They will not accept offers below their floor price (also called the reserve price). To maximize their revenue, they want to set the floor price as close as possible to what the buyer who values the auction item most is willing to pay.[38]

[35] Expert Report of M. Weinberg, ¶18 "In this report, I consider first-price and second-price auctions. In a first price auction, the highest bidder wins the item (allocation) and pays the highest bid (payment). In a second-price auction, the highest bidder wins the item (allocation) but pays the second-highest bid (payment)."

[36] Expert Report of M. Weinberg, ¶47 "In order to understand how auction formats affect bidder strategies, further terminology is needed. A sealed bid single-item auction is truthful if each bidder receives the best possible outcome (given the other bidders' bids) by submitting a bid equal to their own value."

[37] Expert Report of M. Weinberg, ¶¶16-17 ("Auctions can vary in complexity, specificity/concreteness, and format." and "A single-item auction can be held with different rules. This is referred to as the auction format. One such single-item auction format is a sealed bid auction, where the auctioneer solicits a single bid from each bidder and directly decides from these bids to whom to award the item and how much to charge. The term "sealed bid" refers to the fact that each bidder submits a single bid to the auctioneer, in a manner so that other bidders cannot see, and has no further communication with the auctioneer.")

[38] Expert Report of M. Weinberg, fn. 426 ("When there is sufficient data to predict the highest bidder's willingness to pay exactly, the goal is indeed to set the reserve price just below this)"; *id.* at ¶36 ("[R]eserve prices are a useful tool to extract extra revenue from sales…").

63. The seller will run a similar ML model to buyers. Here, it is a given in my ML Abstracted Auction that they know who likely buyers are and have data from past auctions, similar to what the buyers have. In my ML Abstracted Auction, the seller model has access to all data from their own auctions, including all the winning buyers, what they bid, and how much the items have sold for.
64. Given these conditions and these postulates, as with the buyers, if the seller does not have a dataset like this, then they likely will set floors too high or too low. If the floor is too high, then the auction item will not sell. If the floor is too low, they have possibly left money on the table.[39]

3) Auctioneer's role

65. In my ML Abstracted Auction, I do not assume the auctioneer uses any machine learning, because they do not need to. The auctioneer has the following functions.
66. The auctioneer will report buyers' winning bids and clearing prices to the seller. In a 2P auction, if a buyer bids $4 and the auction cleared at $3, the auctioneer reports $3 to the seller, and reports to the buyer that $4 was the winning bid and the auction cleared at $3. If the auction were a 1P auction, the buyer would pay $4.
67. The auctioneer will also take a percentage cut from the clearing price. The cut is fixed and established in advance with the seller.

4) Buyer-Seller-Auctioneer together

68. Now I will demonstrate what happens if the buyer and seller both have ML models based on the same data for valuation and the auctioneer does not use any machine learning. I expect buyer and seller valuation estimates will be similar for the same item.
69. Say for an auction item, the seller sets a floor of $4 based on their model. A buyer put in a bid based on their model for $4.01 and won the auction. The seller's floor price was very close to the winning bid, which means the seller floor and buyer bids are close.

---

[39] Expert Report of M. Weinberg, ¶37 ("For example, if potential homebuyers see the seller setting a low reserve in a first-price auction, they will submit low-ball offers, and the seller will make less money selling their home even if bidders would have been willing to pay a lot more. Setting reserves is a tricky business; set the reserve too low and the seller misses out on extra revenue but set the reserve too high and the seller will miss the sale entirely." (emphasis added)); *id.* at ¶59 ("For example, in the context of a second-price auction, a too high reserve will nullify the entire auction, a too low reserve has no impact, and setting a reserve in the sweet spot between the highest and second-highest bid yields extra revenue.")

70. The auctioneer clears the auction, takes a revenue share, and moves on to the next auction.
71. In the auction setting above, the auctioneer does not use any machine learning to manipulate the auction. Now, I will alter the auction settings such that the auctioneer plays a more active role in changing the rules of the auction and deciding what auction data is shared through "auctioneer manipulation programs."

B. Auctioneer manipulation programs in ML Abstracted Auctions

1) The auctioneer changes the rules: Starting with withholding data

72. The first modification, or manipulation, to the auction behavior the auctioneer manipulation program does is to stop sharing data with the seller. When running a second-price auction, the auctioneer no longer provides the seller with the identity of the buyer; it only provides the clearing price. When it runs a first-price auction, it only provides the winning bid but not the identity of the buyer.
73. The immediate consequence of this modification is that the seller can no longer estimate valuations for each buyer, because they do not know who purchased each item in past auctions. Some of their floors are now much too low for items that are valuable for particular buyers and some of the floors are too high.
74. Buyers can only receive data from auctions that they compete in.[40] However, the auctioneer offers a service to some of the bidders where the bidders can "team up." If they team up, they can use data from every bidder in every auction for their valuations. This new service is called a bidding tool. The bidding tool will conduct an internal auction before sending the bids to the auction. Given these new conditions, buyers are more likely to lose if they do not join the bidding tool, which I will describe shortly.
75. There are many bidding tools in the auction. But the auctioneer favors its own bidding tool. The auctioneer and auctioneer's bidding tool both take percentages of the winning bid, so the auctioneer can earn more if the auctioneer's bidding tool wins.

---

[40] *See* full list of attributes shared with Google RTB buyers here: Authorized buyers. "Real-time Bidding", Accessed on August 19, 2024. https://developers.google.com/authorized-buyers/rtb/openrtb-guide (Bid feedback attribute price: "If the bid won the auction, this is the price paid in your account currency. If the bid participated in the auction but was out-bid, this is the CPM that should have been exceeded in order to win.")

Authorized buyers Help. "Bid data sharing", Accessed on July 24, 2024. https://support.google.com/authorizedbuyers/answer/2696468?hl=en ("Specifically, when a bidder submits a valid bid into the auction, they receive back the minimum value they would have had to bid to win that auction, whether they lost or won.")

76. There are also other auctioneers that sellers could engage to auction their items, and buyers could bid into those auctions.

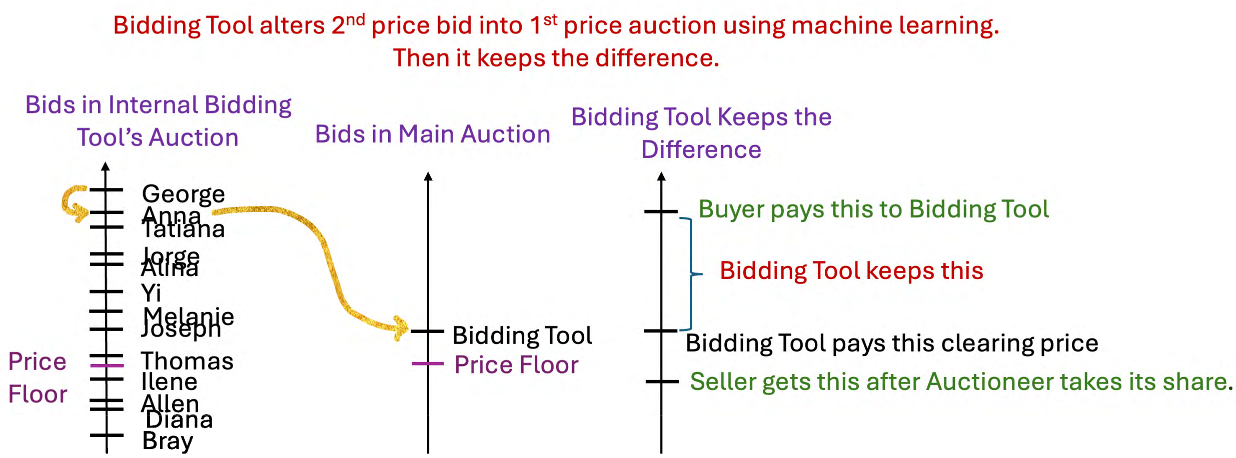
2) <u>Auctioneer's bidding tool starts to adjust bids ("Program B")</u>

77. The auctioneer's bidding tool has data from all past auctions both from the internal auctions and the main auctions. The auctioneer's bidding tool can estimate bids from other bidding tools well. The buyers think the auctioneer's bidding tool places bids on their behalf at their valuations or lower, based on its knowledge, by choosing the lowest possible bid to win the auction.

78. However, the auctioneer's bidding tool does not do that. In this auctioneer manipulation "Program B", the auctioneer's bidding tool has recognized that it can make a profit from its high-quality estimates of its competitor's bids. Without telling anyone, it begins to adjust bids that it places.

79. If the winner of the bidding tool's internal auction has bid much higher than is probably necessary for winning the main auction, it charges the bidder its bid but places a lower bid into the main auction and still almost always wins. It will pocket the difference and do what it likes with it. This is illustrated in Figure 3 below, in the case where the auctioneer's bidding tool's internal auction is a 2P auction and it adjusts the winning second price before bidding into the main auction.

<u>Figure 3: Illustration of "Program B"</u>



80. The buyers and sellers do not have enough data to figure out this is going on.

3) Auctioneer acquires results from other auctions ("Program LL")

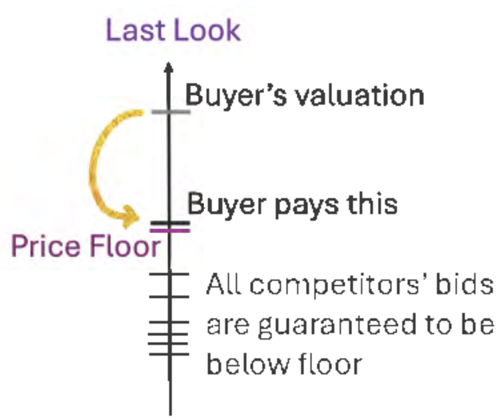
81. In auctioneer manipulation "Program LL," the auctioneer conditions participation for the sellers. In the deal, sellers can participate in the ML Abstracted Auction (otherwise they could not participate), and in return sellers give the auctioneer a number of advantages. These include:

82. The seller will give extra data to the auctioneer. It hands the auctioneer a floor that guarantees all other bids from competing auctions are below that floor.

83. The auctioneer's bidding tool can always bid last.

84. In this case, if the bidding tool's internal auction winner would have been the highest bidder in the main auction, it pays only one penny more than the seller's floor. This is illustrated in Figure 4 below.

Figure 4: Illustration of "Program LL"



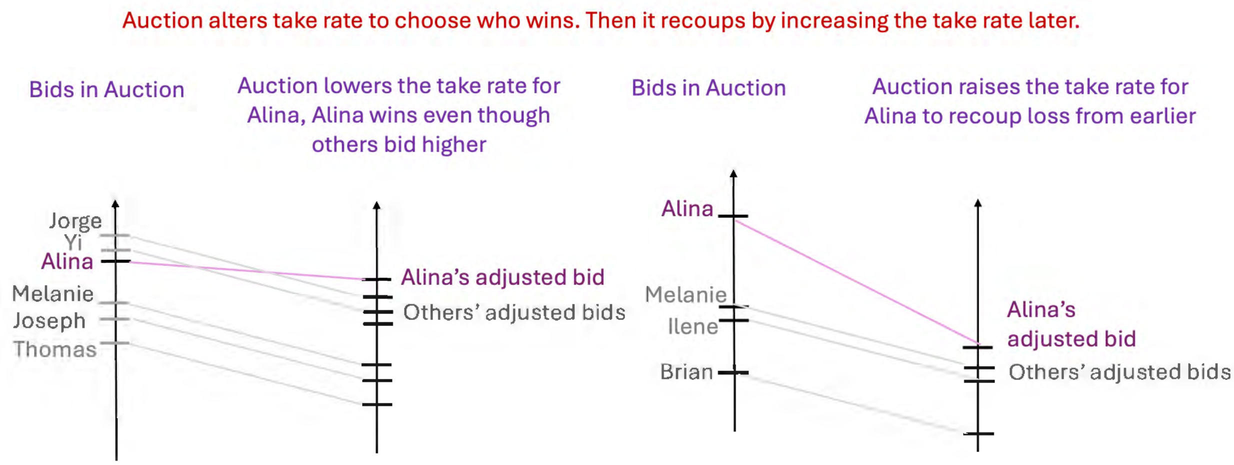
85. Thus, the data from the seller and the opportunity to bid last give the bidding tool's winning buyer a big discount.

4) Auctioneer starts to adjust take rate – ("Program SSD")

86. In auctioneer manipulation "Program SSD", the auctioneer expands the ML Abstracted Auction to include buyers who win auctions run by other auctioneers. The auctioneer has data from all auctions it runs. It has more data than anybody – more than its own sellers and buyers, and more than other buyers because it has all bids that have ever been placed in its auctions and every floor from every seller participating in its auctions.

87. The auctioneer would typically charge a fixed percentage off a clearing price. However, this auctioneer changes its percentage per-auction and per buyer to clear more auctions and choose auction winners. The auctioneer often changes who wins the auction; in fact, when using this program, it forces the top bidder to lose the auction in favor of a lower bidder over [redacted] of the time.[41] The program only applies to some of the bidders. This is illustrated in Figure 5 below.

Figure 5: Illustration of "Program SSD"



88. If the winning bid in the auction is far above the floor, the auctioneer takes a larger percentage than usual. The auctioneer then sets aside the marginal difference and places it into a pool of money. It will pocket this and do what it likes with it.

89. The auctioneer also works with its auctioneer's bidding tool to make sure that both the auctioneer manipulation programs do not happen on the same auctions – for instance, Program B and SSD do not run on the same auctions.

90. The seller model and buyer do not have enough data to figure out that this is going on. Thus, the buyers and sellers will not be able to detect Program SSD, i.e., the auctioneer's adjustments in take rates.

5) Auctioneer starts to adjust floors for some buyers ("Program R")

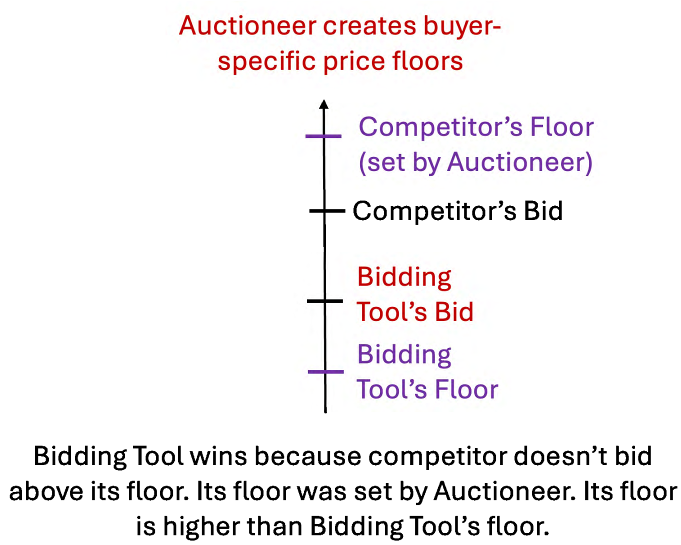
91. As a reminder, in the ML Abstracted Auction, the auctioneer has much more information than the seller, and the seller often sets floor prices too low as a result. The auctioneer also favors its own bidding tool. The auctioneer also allows "other bidding tools to submit bids in the

[41] See Section V.C.3

auctioneer's auctions. In auctioneer manipulation "Program R", the auctioneer sets floor prices for other bidding tools much higher than the seller's floor prices in cases where it is predicted that the seller inadvertently set them too low. It does this without telling the seller. The floor prices can be set per-buyer, so different bidders will be assigned different floor prices.

92. This modification serves two purposes. First, it forces other bidding tools to pay more to win. Second, the auctioneer thus collects more money from these auctions. An illustration is in Figure 6 below, where the winner of the auction actually had a lower bid than its competitor; the competitor did not bid above the personalized auction floor set for it by the auctioneer.

Figure 6: Illustration of "Program R"



93. To do this, the auctioneer leverages its data advantage. It estimates valuations for other bidding tools more accurately than sellers do.

94. The seller does not have enough data to figure out this is going on.

6) Auctioneer interferes in outside deals ("Program E")

95. Buyers and sellers have trouble estimating the valuations of each individual item. It is easier to estimate the total average value of a collection of items. The items have a range of values, but, within the pile, buyers and sellers expect a similar number of high-value items.

96. Thus, buyers and sellers make deals based on estimates of the total value. They agree on a fixed number of items at the total average price. These deals do not involve auctioneers or auctioneer's bidding tool since there is no auction.

97. In addition, with these deals, sellers and buyers can avoid large auctioneer and auctioneer's bidding tool fees; if these items had gone to auction, then the auctioneer and auctioneer's bidding tool would have collected a large fraction of the bidding price.

98. In a normal auction, the auctioneer would not be involved in these outside deals. However, in this case, the auctioneer has access to all of the seller's items, chooses the most valuable ones, and sends them to the auction; this is "Program E." The auctioneer does not disclose its behavior. The seller has no choice in this matter.

7) <u>The auctioneer and its bidding tool throttle the auctioneer manipulation programs</u>

99. The auctioneer and auctioneer's bidding tool do not want sellers and buyers to discover the auctioneer manipulation programs over concern that they might be unhappy with them and/or that they might change their behavior after knowing about them. So, they throttle the programs, meaning that they turn them on and off frequently to inject enough randomness into the bids and floors that the programs are hard to detect by sellers and buyers.

C. The auctioneer and its bidding tool's actions affect buyer and seller models

100. Through auctioneer manipulation programs, the auctioneer and its bidding tool overcharge and put cash into a pool. The auctioneer decides that, instead of keeping this money, they will reinvest it to win even more auctions. More auction wins get it more data to use in future auctions. So it modifies the behavior of the ML Abstracted Auction in the following ways:

101. The auctioneer creates more wins for its bidding tool by using the cash to raise its bid over competitors' bids (it subsidizes the auctioneer's bidding tool.)

102. The auctioneer creates more auction wins when the floor price is set higher than the auctioneer's bidding tool's bid. Whereas earlier, no one would have won the auction, now the auctioneer can simply use their cash from the pool to raise the bid so that auctioneer's bidding tool wins. (Again, the pool subsidizes the auctioneer's bidding tool.)

103. The auctioneer gives some money to the sellers, in the form of raising some of the auctioneer's bidding tools' bids into the auction. This keeps the sellers making enough profit to discourage them from moving to different auctioneers. It's costly for sellers to switch auctioneers.

104. As the auctioneer and the auctioneer's bidding tool begin to implement their programs, the buyer and seller models will struggle to adapt. The auctioneer manipulation programs change the data that the buyer and seller models rely on, leading to worse estimations, predictions and optimizations for valuations, bids and floor prices. This gives the auctioneer more and more of an advantage over sellers and buyers, other auctions, other bidding tools, and the possibility of direct deals.

105. As time goes by, the auctioneer and its bidding tool cement their empire. The auctioneer has an ever-growing dataset of auctions, including both buyer and seller information that no one else has. They use that to keep making better predictions, collecting cash into their pool, and reinvesting it to win more auctions. The sellers cannot create better predictions – they do not know who wins each auction, nor how much they bid, and the bids are subsidized, changing them away from their valuations. This makes setting floor prices even more difficult for the seller, and the cycle continues.

106. While this hypothetical situation was described abstractly, it is a simplified – but very real – version of what happened in practice. Google's online display advertising marketplace revolves around ad auctions conducted in its AdX exchange. These ad auctions are conducted between publishers ("sellers"), who have ad space on their websites ("auction items") that they would like to use for revenue, and advertisers ("buyers"), who place bids on the ad space being auctioned, so that they can have their ads displayed. Google is the auctioneer. Google has engaged in auctioneer manipulation programs that alter auction rules to the benefit of its AdX exchange and Google Ads ("auctioneer's bidding tool") and distort information that buyers and sellers use to optimize reserve prices and bids.

107. Google used auction manipulations to prevent the buyers and sellers in its auctions from optimizing bids and reserve prices. Google relied on machine learning models that leveraged information from all buyers and sellers in its auctions, including both the internal auction of its bidding tool, Google Ads, and its online marketplace, AdX. It used these models to influence the information, bids, and bidding strategies of buyers and sellers. It used ML in ways that are secretive, and sometimes in ways that sent revenue towards one class of buyers

but hurt another class, sent revenue towards some buyers at the expense of sellers, or sent revenue to sellers at the expense of some class of buyers.[42]

108. Google's experts have opined that buyers and sellers can use Google's ML tools to determine the most profitable auction strategies, and that even in the face of Google's auction manipulations, the tools allow buyers and sellers to conduct experiments and set bids and reserve prices correctly.[43] But these experts' opinions ignore the way that Google's auction manipulations interfere with buyers' and sellers' ability to predict and optimize. Without knowledge of when Google makes changes to auction bids or reserve prices, buyers and sellers would find it extremely difficult to optimize, estimate, or predict correctly.

109. For instance, consider what happens when Google manipulated some but not all of buyers' bids. In that case, neither sellers nor buyers were able to rely as much on past auction data to predict what buyers would bid in the future, since they had no way of knowing which auctions

---

[42] The auction manipulations Bernanke, DRS, and RPO, which relied on ML models as I discuss below, were secret for part or all of their duration.

The first iteration of DRS (DRSv1) was secret; see Expert Report of M. Weinberg ¶189 ("DRS was launched in August without announcing it to publishers or advertisers.")

The second version of DRS (DRSv2) was disclosed, but with caveats. See Expert Report of J. Gans ¶779 ("Google chose not to make publishers aware of the take rates for each transaction. This was an intentional choice by Google. Instead, in later implementations, publishers were given an opportunity to opt out of DRS but without proper information to make that choice"); *see also* Expert Report of M. Weinberg ¶197 ("DRSv2 was launched in the second half of 2016. Google announced DRSv2 when it was launched. The publishers were allowed to opt out of DRSv2, however, if they did, Google turned off DRSv1 for these publishers as well. Advertisers and ad buying tools could not opt out of DRSv2.")

On the secrecy of RPO, see Section V.D.2 below. In addition, as I discuss in Section V.F, these programs frequently throttled on and off in ways that buyers and sellers would not have known about.

Regarding behavior that favored some buyers over others, see Expert Report of M. Weinberg ¶232 ("I show that that Projects Bernanke and Global Bernanke . . . led to a lower win rate for non-GDN ad buying tools and advertisers that use those ad buying tools.")

Regarding behavior that tended to favor some buyers over sellers, see discussion of DRS in Expert Report of M. Weinberg ¶12.b.ii ("Dynamic Revenue Sharing version 2 (DRSv2), in comparison to both no DRS and DRSv1, decreased advertiser payoff . . . and may also have decreased publisher revenue.")

Regarding behavior that tended to favored sellers over buyers, see discussion of RPO in Expert Report of M. Weinberg ¶272 ("I demonstrate that RPO leads to higher revenue for Google's ad exchange AdX and explain the mechanisms through which it leads to lower payoff to advertisers and could lead to lower revenue for some publishers.")

[43] Expert Report of P. Milgrom, ¶32 ("And even for those programs for which details are not disclosed at all, advertisers' and publishers' routine data analysis and experimentation with bids and floor prices are typically sufficient for them to identify optimal strategies.")

Expert Report of A. Ghose, ¶158 ("In addition to relying on their ad tech providers' floor price optimizations, Publishers can also test the outcomes of different floors to determine how best to optimize their revenue.")

Expert Report of N. Wiggins, ¶86 ("When one publisher was losing auctions due to high price floors, Google reached out proactively to suggest that the publisher use Google's automated tool for optimizing price floors or that it test alternative price floors itself using Google's manual experiments tool.")

have been affected by Google's auction manipulations. Similarly, when Google altered floor prices of sellers (without their knowledge) that applied to some bidders but not others, sellers would have trouble figuring out which auctions were affected and how to adjust their floor prices. Buyers whose floors were raised closer to their valuations would find it harder to optimize bids than others whose floors were not affected. Throttling also introduces randomness into bids and floors such that it is difficult for buyers and sellers to detect this behavior. In the following sections, I will show how, in the real-world, Google's auction manipulations make it difficult for advertisers and publishers to experiment to identify optimal strategies.

IV. BUYERS AND SELLERS DO NOT HAVE ACCESS TO THE SUBSTANTIAL FRESH DATA THAT GOOGLE ITSELF REQUIRES TO OPERATE ML MODELS RELATED TO AUCTIONS

110. Prof. Ghose's opinion that "[w]ith digital advertising, advertisers have access to large amounts of data, sophisticated analytics tools, and real-time insights, which enable them to make more informed decisions about budget allocation" is exaggerated.[44] Prof. Ghose downplays the massive data advantage Google has historically retained—and still has—over individual advertisers and publishers.

111. Throughout the development and deployment of its ML models and auction manipulations, Google had distinct data advantages compared to individual buyers and sellers, which benefited its ML models for valuations, prediction of competitors' bids, and floor optimization. As I discussed in Section II.B access to relevant and fresh data is critical for the performance of machine learning models.[45]As I will discuss, buyers and sellers had access to data only for

[44] Expert Report of A. Ghose, ¶105 ("With digital advertising, advertisers have access to large amounts of data, sophisticated analytics tools, and real-time insights, which enable them to make more informed decisions about budget allocation."); *id.* at ¶109 ("These data provide specific and actionable insights, allowing advertisers to adapt quickly and optimize their campaigns. Traditional analytical methods, as well as machine learning and A/B testing, are increasingly used to analyze digital advertising data. Digital advertising also enables more precise targeting, ensuring that ads reach specific, defined audiences.").

[45] Budach, Lukas, et al. "The effects of data quality on machine learning performance." arXiv preprint arXiv:2207.14529 (2022). ("Modern artificial intelligence (AI) applications require large quantities of training and test data. This need creates critical challenges not only concerning the availability of such data, but also regarding its quality. For example, incomplete, erroneous or inappropriate training data can lead to unreliable models that produce ultimately poor decisions. Trustworthy AI applications require high-quality training and test data along many dimensions, such as accuracy, completeness, consistency, and uniformity.")

Prapas, I., Derakhshan, B., Mahdiraji, A.R. et al. Continuous Training and Deployment of Deep Learning Models. Datenbank Spektrum 21, 203–212 (2021). https://doi.org/10.1007/s13222-021-00386-8. ("Several modern applications like financial and recommender systems require models that are constantly updated with fresh data.")

auctions they participated in (and only minimal anonymized data for their own auctions), whereas Google had access to data about *all* the auctions Google ran, as well as detailed data about the impressions being sold and the business goals of both buyers and sellers, including Google's third-party cookie data and data from Google's Owned and Operated Properties.[46] In other words, Google's enormous scale and access to detailed data conferred an insurmountable advantage. And, as I discuss further below in Section V, Google further widened its data advantage by skewing buyers' and sellers' data through Google's auction manipulations. Whereas individual buyers or sellers could only conduct experiments or A/B testing across their own limited data, Google's display ad business could leverage all the data it had access to, as well as knowledge of its own auction manipulations.

112. The sections below detail Google's unique data advantage derived from different data sources, which manifests in its machine learning programs.

A. Data advantage from past auctions

113. [redacted][47] Google benefits from economies of scale for data collected from past auctions. Collecting data on the winning bid is particularly valuable, since the associated data could include whether the user clicked on the bidder's advertisement or whether the user made a

---

[46] Google's ability to collect first- and third-party data in the United States is dominant. Publicly accessible data show that Google indeed owns the largest number of data trackers or cookies on the web.

Spread Privacy. "DuckDuckGo Tracker Radar Exposes Hidden Tracking", Accessed on September 9, 2024. https://spreadprivacy.com/duckduckgo-tracker-radar/ ("[T]he data set shows Google-owned trackers are on over 85% of the top 50K sites")

Google Search is the most popular search engine, with 88% market share. statcounter, "Search Engine Market Share United States of America," https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america. Accessed September 7, 2024.

YouTube is the most popular video streaming service, with 81% market share. Auxier, B., Anderson, M., "Social Media Use in 2021," https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/. Accessed September 7, 2024.

Google Maps is the most popular map service, with 71% market share. Statista, "Most popular mapping apps in the United States as of April 2018, by reach," https://www.statista.com/statistics/865419/most-popular-us-mapping-apps-ranked-by-reach/. Accessed September 7, 2024.

Google tracks users on 78% of websites. DuckDuckGo, "tracker-radar/entities/Google LLC.json," https://github.com/duckduckgo/tracker-radar/blob/main/entities/Google%20LLC.json. Accessed September 7, 2024.

[47] GOOG-NE-13197548 at '549, "AdX Overview: For Nooglers" (April 1, 2014) - Internal Presentation on AdX for new employees. [redacted])

purchase.[48] Google's access to detailed bid data from its past internal auctions and AdX auctions place it in a unique position to train its machine learning models to optimize bids and predict the value of an impression when shown to a user.[49]

1) Bid data from Google Ad's internal auctions

114. Data recorded from past auctions includes a variety of bid-level information, as described by a research team:

> [T]he user (e.g., the user segmentation), advertiser (e.g., the creative format and size), publisher (e.g., the auction reserve price, ad slot, page domain, and URL) and the context (e.g. the time, region, the browser, and the operation system). For each bid request, there is an auction-winning bid price, and the user feedback (click, conversion) is recorded if the advertiser won the auction.[50]

115. Google has access to this detailed bid-level data for all queries transacted in their internal auctions in Google Ads. I have seen no evidence that Google Ads buyers get any bid data except for their own bids and clearing price information for the internal auction. I have seen no evidence that sellers have access to any internal auction bid data at all.

2) Bid data from AdX auctions

116. Google collects detailed bid-level data from all AdX auctions, which includes auction results, such as winning bids and payment information, and user activities, such as their past Google search queries and ad views and user interactions resulting from an ad (clicks and

[48] Google Ads Help. "Evaluate ad performance on the Display Network", Accessed on September 8, 2024, *cited in* Expert Report of J. Gans, ¶49 ("The advertiser can also then track the subsequent actions taken by the user (e.g. whether they clicked on the ad, or subsequently made a purchase), which further helps them identify the types of users that are most valuable.")

See also Zhang, W. Yuan, S., Wang, J. "Optimal Real-Time Bidding for Display Advertising" KDD '14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. p.1082 "For each bid request, there is an auction winning bid price, and the user feedback (click, conversion) is recorded if the advertiser won the auction."

[49] Zhang, W. Yuan, S., Wang, J. "Optimal Real-Time Bidding for Display Advertising" KDD '14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ("This bid calculation (see the bidding engine in Figure 1) is the most important problem for a DSP. The solution to this problem is referred to as a bidding strategy. In pure second price auctions for strategic competitors, theoretically the dominant strategy for advertisers is truth-telling: to bid their private values. When facing a bid request, a DSP will evaluate the value of the impression i.e. to estimate the click-through/conversion rate (CTR/CVR) and multiply it by the value of a click/conversion. Many advertisers simply set this value as their bid and keep using it throughout a campaign's lifetime.")

[50] Zhang, W. Yuan, S., Wang, J. "Optimal Real-Time Bidding for Display Advertising" KDD '14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. p.1081-1082.

views).[51] In first price auctions, buyers (which are mostly bidding tools) only have access to the minimum-bid-to-win information.[52] Sellers receive the clearing price of the auction, as well as the Bid Data Transfer (Bid DT) files, which contain the collection of AdX bids.[53] The bidders in AdX auctions (who are mostly bidding tools) are anonymized.[54] That is, *sellers do not know who won the auction.* Internal documents also suggest that AdX bidders can opt out, making their bids unavailable to sellers; thus, the sellers *may not even receive the clearing price of the auction.* In addition, in 2019, Google started [redacted] so that they are no longer able to link the users to the bids; *thus, Google prevented sellers from knowing who they are serving the ads to.*[55]

117. Google can combine all data from sellers and buyers to run experiments, an advantage over sellers and buyers who can only use the limited and anonymized data from auctions they participate in. For instance, [redacted], former Tech Lead Manager at Google, testified that Google knew that it was not possible for Google or sellers to optimize per-buyer reserve floors with the data available to a single seller.[56] [redacted]

---

[51] Expert Report of J. Hochstetler, fn. 658 ("Google data production, samples of Google Ad Manager (GAM) log-level bid data produced on May 12, 2023. [redacted])

[52] *See* full list of attributes shared with Google RTB buyers here: Authorized buyers. "Real-time Bidding", Accessed on August 19, 2024. https://developers.google.com/authorized-buyers/rtb/openrtb-guide (Bid feedback attribute price: "If the bid won the auction, this is the price paid in your account currency. If the bid participated in the auction but was out-bid, this is the CPM that should have been exceeded in order to win.")

Authorized buyers Help. "Bid data sharing", Accessed on July 24, 2024. https://support.google.com/authorizedbuyers/answer/2696468?hl=en ("Specifically, when a bidder submits a valid bid into the auction, they receive back the minimum value they would have had to bid to win that auction, whether they lost or won.")

[53] Expert Report of P. Pathak, ¶150 ("Google provides two types of data files to DFP publishers: Data Transfer files, which include Header Bidding bids, and a Bid Data Transfer file, which includes AdX and Exchange bidding bids, commonly referred to as "DT" files.")

[54] GOOG-NE-04599495 at '495, "+External position – 1P Bid Data Transfer Balancing transparency to publishers, with protecting buyer data and user privacy" (No date)- Internal Google document discussing Bid data transfer files ([redacted])

[55] GOOG-NE-06879156 at '159, "1p Auction - Bid Data Transfer Roll-out plan" (July 24, 2019) - Internal Google presentation [redacted])

[56] Deposition of [redacted] (Former Tech Lead Manager, Google), 113:2-113:4, May 23, 2024 ("With per buyer, that's getting inside the mechanism of the auction itself. And a publisher can't do that on their own.")

[redacted].[57] Thus, while Google has access to detailed auction data that it can use for ML models, sellers and buyers do not have the same access.[58]

B. Data advantage from detailed user data

118. Google has access to large amounts of fine-grained user data it collects through its third-party cookies and its Owned and Operated properties as discussed below.

1) Data from third-party cookies

119. One of the main ways Google collects user data is through cookies, or small files of information, that are created when a person visits a website.59 Cookies typically store user data, such as web surfing and personalization preferences, with an ID unique to the user and device.60 This data stored in cookies can then be used to personalize the user's experience and track user data over time.61 Third-party cookies are placed on a website by someone other than the website owner. Third-party cookies play a crucial role in online data collection because they allow the entity that placed the cookie (the cookie's "owner," here Google) to

---

[57] Deposition of [redacted] (Former Tech Lead Manager, Google), 142:4-142:8, May 23, 2024. ("Q If we can go to the next e-mail. It's from [redacted], and he writes: [redacted]

GOOG-AT-MDL-019653406 at '418, "Re: [Update] [redacted] - RPO from DFP" (July 7, 2014) - Internal Google email thread ([redacted] discussing DFP and AdX merger [redacted]

[58] GOOG-DOJ-04937154 at '159, "Optimized pricing in the Open Auction Comms" (Mar. 23, 2018) – Google communications document on optimized pricing in the open auction ("AdX optimized pricing uses all event level AdX data. To price as intelligently as possible, all event level data from past Open Auctions may be considered (subject to contract terms). This means a buyer's historic bids may be used as part of the prediction model to set a price for them. Please note inputs to optimized pricing remain confidential to the buyer per the contract.")

[59] IETF. "Cookies: HTTP State Management Mechanism", Accessed September 9, 2024 https://www.ietf.org/archive/id/draft-ietf-httpbis-rfc6265bis-07.html#name-acknowledgements ("Using the Set-Cookie header field, an HTTP server can pass name/value pairs and associated metadata (called cookies) to a user agent. When the user agent makes subsequent requests to the server, the user agent uses the metadata and other information to determine whether to return the name/value pairs in the Cookie header.")

[60] TechTarget. "Third-party cookie" (May 2023), Accessed August 30, 2024. https://www.techtarget.com/whatis/definition/third-party-cookie ("Cookies typically store user data such as web surfing and personalization preferences and tracking information.")

[61] IETF. "Cookies: HTTP State Management Mechanism", Accessed September 9, 2024 https://www.ietf.org/archive/id/draft-ietf-httpbis-rfc6265bis-07.html#name-acknowledgements ("Cookies are often criticized for letting servers track users. For example, a number of 'web analytics' companies use cookies to recognize when a user returns to a web site or visits another web site. Although cookies are not the only mechanism servers can use to track users across HTTP requests, cookies facilitate tracking because they are persistent across user agent sessions and can be shared between hosts.")

directly track a user across different websites where its cookies are placed.[62] Third-party cookies are often used by ad servers to track user behavior across several websites where ad servers provide ads. Sellers who use Google's ad server also have Google's third-party cookies placed on their websites. Google uses these cookies to track user behavior on these websites and collect granular information on user demographics, interests, and geolocations.[63]

120. Publicly accessible data show that Google owns the largest number of data trackers or cookies on the web. For instance, DuckDuckGo's Tracker Radar measurements in 2020 indicated that "Google-owned trackers are on over 85% of the top 50K sites."[64] Google has access to fine-grained user data, amounting to over [redacted].[65] In sum, Google continually collects a significant amount of data from a number of people [redacted] which Google can use to train its ML models.

2) Data advantage from Google's Owned and Operated (O&O) properties

121. In deposition testimony, former Google employee [redacted] explained why Google has better data access than sellers and buyers do; it is because Google has highly detailed data on people who use Google across Google Owned and Operated (O&O) properties. 66 As [redacted] noted, "by virtue of having [unique identifiers] on

---

[62] TechTarget. "Third-party cookie" (May 2023), Accessed August 30, 2024. https://www.techtarget.com/whatis/definition/third-party-cookie ("[T]hird-party cookies track users across several websites, providing a more comprehensive picture of user behavior.")

[63] GOOG-AT-MDL-B-005177744 at '747, "Field Guide to Display Ads" (January 27, 2023) - Internal Google guide for new engineers covering basics of display ads ("GDN also tracks user behavior on AdSense publisher sites to create a profile of interests. Profile data enables advertisers to target user interest categories, such as travel, fashion, and so on. In addition, Google buys user data from third parties to enhance user-interest targeting.")

[64] "DuckDuckGo Tracker Radar Exposes Hidden Tracking", Accessed on September 9, 2024. https://spreadprivacy.com/duckduckgo-tracker-radar/ ("[T]he data set shows Google-owned trackers are on over 85% of the top 50K sites.")

[65] GOOG-DOJ-AT-02504615 at '616, '[redacted] (No Date) - Internal Google document discussing [redacted] and leveraging user data to train prediction models. [redacted]

[66] Expert Report of J. Gans, ¶56 ("Google itself is a publisher that owns and operates multiple web and mobile properties including: Google Maps (online map service), Gmail (email), Google Search (search engine), Android (mobile operating system), Google Chrome (web browser), and YouTube (video-sharing site). Google's vast array of properties allows Google to collect data and track users, and monetize various digital advertising channels, such as search ads on Google Search and video ads on YouTube.")

properties…like YouTube and [M]aps, you could then recognize the user across lots of properties and across lots of devices. So you could identify that like the user on a -- this mobile phone and that mobile phone are the same user, or this mobile phone and that desktop are the same user, or this search request is also that user."[67] She added that small buyers would not have access to such types of detailed data.[68]

122. Google's rich O&O data (collected through products such as YouTube, Google Play, Google Search, Google Maps, etc.) provides Google tools with a data advantage that is unavailable to third parties, including buyers, sellers, and other ad exchanges.[69] Google supplements data collected through third-party cookies with its O&O data to improve targeting for its ad buying tools.[70] The user activity data from across Google's suite of products and

---

GOOG-AT-MDL-018520983 at '984, "Owned &Operated Ad Creation" (No Date) - Internal Google document on Ads in Google Owned and Operated properties, ("• Ad creation for a new campaign type for Google Owned & Operated (O&O) properties...• Gmail| YouTube Display Ads • Video ,Apps and Search retain their existing buying flows. •Home Feed, Image Search, News and other apps later in 2019).")

[67] Deposition of [redacted] (Former Tech Lead Manager, Google), 223:7-223:14, May 23, 2024 ("And by virtue of having this on properties that had a very large like YouTube and maps, you could then recognize the user across lots of properties and across lots of devices. So you could identify that like the user on a -- this mobile phone and that mobile phone are the same user, or this mobile phone and that desktop are the same user, or this search request is also that user.")

[68] Deposition of [redacted] (Former Tech Lead Manager, Google), May 23, 2024, 224:4-224:6, ("And competitors, as you testified, simply don't have the scale like Google to engage in this type of conduct?); *id.* at 224:12-224:15 ("Q. Or, for example, smaller -- in the display ads context, smaller Ad Exchanges, add servers and buying tools? A. They definitely wouldn't, because they wouldn't have the properties to get the information from.")

[69] Google Search is the most popular search engine, with 88% market share. statcounter, "Search Engine Market Share United States of America," https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america. Accessed September 7, 2024.

YouTube is the most popular video streaming service, with 81% market share. Auxier, B., Anderson, M., "Social Media Use in 2021," https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/. Accessed September 7, 2024.

Google Maps is the most popular map service, with 71% market share. Statista, "Most popular mapping apps in the United States as of April 2018, by reach," https://www.statista.com/statistics/865419/most-popular-us-mapping-apps-ranked-by-reach/. Accessed September 7, 2024

Google tracks users on 78% of websites. DuckDuckGo, "tracker-radar/entities/Google LLC.json," https://github.com/duckduckgo/tracker-radar/blob/main/entities/Google%20LLC.json. Accessed September 7, 2024.

[70] Google's Chrome browser is the most popular web browser, with 52% market share in 2024. Statista, "Market share of leading internet browsers in the United States and worldwide as of August 2024", https://www.statista.com/statistics/276738/worldwide-and-us-market-share-of-leading-internet-browsers. Accessed September 7, 2024.

Google's ability to collect first and third-party data in the United States is dominant. Google Search is the most popular search engine, with 88% market share. statcounter, "Search Engine Market Share United States of America," https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america. Accessed September 7, 2024.

YouTube is the most popular video streaming service, with 81% market share. Auxier, B., Anderson, M., "Social Media Use in 2021," https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/. Accessed September 7, 2024.

Google Maps is the most popular map service, with 71% market share. Statista, "Most popular mapping apps in the United States

services are extremely valuable signals (i.e., input features) for personalized ad targeting that is not available to any other third-party tool.[71]

123. Further starting in 2016, Google was able to combine their data advantage from third-party cookies and O&O data through Project Narnia 2.0.[72] Through the program, Google was able to link people's data associated with their Google account to data collected by its third-party cookies for their ad targeting models. The Google O&O data associated with a user's Google account includes but is not limited to:[73]

a. User's search history on Google Search;
b. User's Chrome browsing history;
c. User's YouTube view history;
d. User's app usages on Android; and
e. User's location history, including physical store visits.

124. When Google supplemented third-party cookie data with Google's data from O&O products, it subsequently created a single view of an individual's data and identity for use throughout Google's display advertising tools and other O&O products. 74 For example, Google can link a person's past Google Search history to show them highly customized ads when they visit a publisher website.

C. Google leverages these and other data advantages in their ML programs used in their display advertising tools

125. Google's display advertising tools use machine learning for several purposes, including 1) to predict whether someone browsing a seller's website would be interested in seeing a buyer's

as of April 2018, by reach," https://www.statista.com/statistics/865419/most-popular-us-mapping-apps-ranked-by-reach/. Accessed September 7, 2024

Google tracks users on 78% of websites. DuckDuckGo, "tracker-radar/entities/Google LLC.json," https://github.com/duckduckgo/tracker-radar/blob/main/entities/Google%20LLC.json. Accessed September 7, 2024.

[71] In an internal communication, Google confirmed the distinct value of "user intent" information obtained from Search and Chrome activity, noting that "proactive [search] query is such a strong signal of user intent," as a user is much more likely to make a purchase decision after having searched for it. GOOG-NE-06230051 at '052, "In-market Audiences on GDN harmful" (No Date) – Internal Google communication on use of search data as a signal for audience targeting.

[72] GOOG-AT-MDL-007418936 at '938, "Narnia2 Overview" (2016-12-14) – Internal Google presentation on Narnia 2.

[73] GOOG-DOJ-28501885 at '892-'895, "Narnia2 Overview (2015-01-20) – Internal Google presentation on Narnia 2.

[74] GOOG-AT-MDL-016354429 at '429, "Narnia 2 Gaia keyed Serving End State Design (Tinman)" (2017-06-12) – Internal Google design document on Narnia 2.0 and GKS implementation.

ad, 2) to forecast how that person will interact with the ad, and subsequently 3) to estimate the value of having the person see the ad. This is necessary because buyers pay Google to reach individuals with ads that may be effective, and sellers want to set a floor price, which is the minimum amount they are willing to accept from buyers for the item being auctioned. Buyers benefit when their ads cause people to make purchases, but buyers do not benefit if the ads are too costly. Thus, models of click-through-rate are important in creating valuations of impressions. Bidding strategy is important in the ML context as well. A buyer may seek to predict what other buyers will bid in order to outbid them. Sellers also need to optimize floor prices. If there are not many buyers, the floor price is important, and the seller might want to predict the bids in order to set the floor just below the top bid(s).[75] In the following sections, I will discuss how Google employs ML in its advertising tools for different predictions, estimations, and optimization and how these models benefit from Google's data advantages.

126. **Targeting:** Google uses "ML to classify audience based on millions of pieces of data" to show them ads based on their interests.[76] These complex models can "pick up on non-obvious correlations by finding data patterns from users."[77] The models are trained on large quantities of contextual signals from people visiting the publisher website.[78] For example, visitors to Nike.com may be much more likely to be interested in purchasing athletic shoes than people who do not visit Nike.com. Google has access to large amounts of user data through their third-party cookies and O&O data, as discussed earlier, which allows them to "aggregate learnings from user's journey online, and then using [m]achine [l]earning to analyze patterns in their

---

[75] *See* Section III for more detailed information on how the auctions work.

[76] GOOG-AT-MDL-018621551 at '551, "Machine Learning (ML) for Display Audience Targeting" (May 20, 2019) - Internal doc to help Google learn how Google uses ML to Enhance its Display Audience Targeting types [redacted]

[77] GOOG-AT-MDL-018621551 at '551, "Machine Learning (ML) for Display Audience Targeting" (May 20, 2019) - Internal doc to help Google learn how Google uses ML to Enhance its Display Audience Targeting types [redacted]

[78] GOOG-AT-MDL-018621551 at '555, "Machine Learning (ML) for Display Audience Targeting" (May 20, 2019) - Internal doc to help Google learn how Google uses ML to Enhance its Display Audience Targeting types ("One way we can make Google Audiences even more relevant and helpful is [redacted]

online journeys."[79] [redacted].[81] Targeting data is an input into estimating the value of an impressions when shown to a user.

127. **Valuation models:** Valuation models allow Google to estimate how well an ad will perform once shown to a person, i.e., the chances that a person will click on an ad (predicted Click-Through Rate, or pCTR) or subsequently purchase a product from the ad (predicted Conversion Rate, or pCVR).[82] These predictions are critical for Google as they inform the value of an ad impression to the buyer. Valuation models are key inputs to Google's bid optimization models, helping Google to evaluate the return on investment (ROI) on a buyer's bid based on the expected performance of an impression. Google can estimate the valuation of an impression to a buyer using detailed data from the user, including browsing history, location, recent Google searches, and demographic data like age, gender, daily clicks.[83]

128. Importantly, to estimate the value of an ad, that ad (or a similar one) needs to be shown to a user to observe their reaction. That is, the more the ad buyer has won auctions in the past, the more auction win data is available to estimate the value of an ad.

---

[79] GOOG-DOJ-AT-02504615 at '616, "[redacted]" (No Date) - Internal Google document discussing [redacted] and leveraging user data to train prediction models. [redacted] GOOG-AT-MDL-018621551 at '555, "Machine Learning (ML) for Display Audience Targeting" (May 20, 2019) - Internal doc to help Google learn how Google uses ML to Enhance its Display Audience Targeting types.

[80] GOOG-AT-MDL-012760228 at '228-'229, "Expanded Interest Based Targeting in GDN [redacted] (No Date) - Internal Google document [redacted]

[81] Ibid.

[82] GOOG-AT-MDL-B-005785771 at '776-779, "Machine Learning in the Google Display Network" (No Date) - Internal doc on how ML helps with performance prediction, including why it's so hard to get the ML models right, "Selected applications of machine learning: Predict Click Through Rate; Quality of Clicks".

[83] GOOG-AT-MDL-001283820 at '825, "Overview of CTR prediction for GDN ads" (No Date) - Internal Google presentation providing an overview of click through rate prediction in GDN [redacted]

129. As discussed, the seller does not even know who is bidding in the auction or who has won past auctions. Sellers do not have access to even a single past bid from any bidder into any internal auction, and so cannot estimate the value of an impression to any Google Ads buyer.

130. **Bid Optimization:** Google's advertising buy-side tools offer buyers auto-bidding options that allow Google to place bids on their behalf.[84] Google uses ML models to estimate the optimal bid to place for an impression.[85] These bid optimization models are trained on signals from Google's ML models for ad performance prediction and granular bid-level data from historical auctions in AdX including auction outcome (for example, clearing price).[86] Google analyses millions of signals to determine the optimal bids for the right impression, both for its internal Google Ads auction and the AdX auction.[87]

131. Google can predict a non-Google-buyer's bid using all of its past bid data into any of the AdX auctions, as well as the detailed data from the user. In contrast, *the seller does not even know who is bidding in the AdX auction or who has won past auctions. Thus, the seller cannot predict any bids for any bidder.*

132. **Floor Optimization:** Google also uses ML models to set floor prices (also called "reserve prices") for sellers. These models were trained on data from historical auctions, including the bidder information and clearing price for the auction. Google can optimize price floors using its detailed data about users to estimate the value of the current impression to the current user.

---

[84] Google Support. "About Smart Bidding", Accessed on September 4, 2024. https://support.google.com/google-ads/answer/7065882?hl=en ("In bidding, machine learning algorithms train on data at a vast scale to help you make more accurate predictions across your account about how different bid amounts might impact conversions or conversion value. These algorithms factor in a wider range of parameters that impact performance than a single person or team could compute.")

[85] GOOG-DOJ-03151263 at '273, "DBM Optimization" (No Date) – Internal Google presentation discussing integration of GDN optimization strategies to DBM "(Bid Optimization: An algorithm that learns a model of the market of available impressions and chooses a bid for each auction to optimally satisfy the advertiser goal")

[86] GOOG-DOJ-03151263 at '273, "DBM Optimization" (No Date) – Internal Google presentation discussing integration of GDN optimization strategies to DBM ("Bid Optimization: An algorithm that learns a model of the market of available impressions and chooses a bid for each auction to optimally satisfy the advertiser goal")

GOOG-AT-MDL-004434946 at '946, [redacted] (2020-07) - Internal Google document providing an overview of [redacted] model [redacted]

[87] GOOG-AT-MDL-006597606 at '616, "Machine Learning + Google Ads: A Guide on how to talk about ML in ads with customers" (January 1, 2017) - Internal doc to help Google employees discus ML in ads with customers in general ("Smart Bidding in AW, DS, & GDN. How: factors in millions of signals to determine the optimal bid and continually refines models of conversion performance at different bid levels to get more from marketing budgets.")

When Google is trying to set the floor for a newly arriving impression, it can find similar past auctions, for the same ad slot and for the same publisher, including all buyers' bids for those auctions.[88] It would use this data to create price floors that are specific to the buyer, the ad slot, and the specific user. Google emphasizes the need for very fresh data points, as recent as [redacted] for these models to optimally estimate price floors.[89]

133. In contrast, the seller does not know who has bid in any of the auctions. The seller sees the user's past visits to their own site but can't track the user across multiple sites. The seller does not have access to the O&O data. If the user has recently been to another site selling the same type of product, the seller does not receive that information, even though it could be very valuable for setting price floors.

134. **Testing and Experimentation:** As part of developing ML models, Google runs experiments to test the effects of new programs. Namely, Google's experimentation starts from an initial model with baseline performance, changes the parameters or features of the model, and records the progress of the experiments.[90] Measuring the impact of a change to a bidding ML model is one example of A/B testing, where one set of results is compared to a baseline.[91] Google maintains the statistical results of all its experiments [redacted], allowing easy tracking and analysis for the various experiments its ML models go through.[92] Google experiments have access to real-time data for

---

[88] GOOG-NE-13204729 at '746, "AdX Dynamic Price" (August 17, 2015) - Internal Google presentation discussing Reserve Price Optimization design. [redacted]

[89] GOOG-NE-13204729 at '746, "AdX Dynamic Price" (August 17, 2015) - Internal Google presentation discussing Reserve Price Optimization design. [redacted]

[90] Google Developers. "Experiments" (July 17, 2024), Accessed on September 6, 2024. https://developers.google.com/machine-learning/managing-ml-projects/experiments ("Start by establishing a baseline metric. The baseline acts as a measuring stick to compare experiments against.… Make only a single, small change at a time, for example, to the hyperparameters, architecture, or features. If the change improves the model, that model's metrics become the new baseline to compare future experiments again.")

[91] GOOG-AT-MDL-019528391 at '396, "Audience Signals NDA" (No Date) - Internal Google NDA disclosure on use of demo signals in response to fair lending concerns [redacted]

[92] GOOG-AT-MDL-004080757 at '771, [redacted] (April 12, 2019) - Internal Google presentation on experiment platforms [redacted]

about every ad currently running.[93] When Google experimented, it could direct anywhere from [redacted] of its traffic to experiments—on days when it was handling billions of auctions a day.[94] As I will discuss in Section VI, these kinds of experiments would be difficult for buyers or sellers to conduct because of scale and cost (and many experiments would be impossible since the data required is not available to them). In fact, Google itself recommended that for each A/B test, a buyer direct 50% of their traffic for 2 weeks to the test. This could be extremely costly.[95]

D. Google provided buyers and sellers with limited access to data

135. As I discussed, despite having access to large amounts of auction and user data from sources detailed above, Google only provided individual buyers and sellers with a limited amount of data related to each party's own successful bids in particular auctions.[96] As discussed, I have not seen evidence that Google provided non-Google buyers and sellers with data about the other bids made on Google auctions, nor information on auctions where the participants bid but were below the publisher's floor price, nor have I seen evidence that sellers received all data from the internal Google Ads auction, or information about who won the AdX auction, or user data from that auction. By contrast, Google had information about all bids and

[93] GOOG-AT-MDL-004080757 at '771, [redacted] (April 12, 2019) - Internal Google presentation on experiment platforms [redacted]

[94] Deposition of [redacted] (Software Engineer, Google), 106:6-106:12, 106:22-106:23, 107:4-107:6, 107:20-107:24, May 1, 2024, ("Q. Okay. And what is the range of percentages of traffic or auctions that still qualifies as—that you've seen as an experiment? [redacted]

[95] GOOG-AT-MDL-B-005714587 at '588, "Google Ad Manager: Making price floors work for you" (No Date) - External Google communication on UPR ("For now, if you are interested in testing the efficacy of optimized floors on your network, we encourage you to run a manual experiment on an existing UPR with a fixed or Target CPM floor. To get the most reliable results, we would recommend running A/B experiments on at least 50% of traffic for a minimum of two weeks.")

[96] GOOG-DOJ-04937154 at '159, "Optimized pricing in the Open Auction Comms" (Mar. 23, 2018) – Google communications document on optimized pricing in the open auction. ("Q. Will I have any way of knowing when the floor price I pay is set by this optimization? A: Not at this time. Note also that you don't know today whether the floor price you pay was set by a competing bid, a contending booking from DFP, or a publisher-set floor price. So we are not reducing transparency.")

all auction outcomes that it used for its own models.[97] In the absence of these types of data, participants would have had more difficulty estimating as well as Google did.

136. To provide more history, as in the ML Abstracted Auction, Google withheld data. Specifically, in 2019, Google reduced the information sellers receive by unlinking bid data transfer files (Bid DT files) from other data transfer (DT files) shared with sellers.[98] This redaction prevented sellers from linking users to their bids and prevented sellers from determining what advertisers were willing to pay for impressions. Information from bid data transfer files had previously been used heavily by sellers to set floor prices and run experiments under different floor prices, and the redaction "reduce[d] the usefulness of Bid DT files for publishers to enable pricing strategy decisions."[99] In an internal communication discussing a response from a seller—[redacted]—to changes to bid data transfer files, Google admitted that "[these] changes are to be disruptive to [redacted] yield management practices and their measurement of success from the internal ad network."[100] Sellers were unhappy with these redactions, since Google further limited the useful information sellers had available to optimize floors and conduct experiments.[101]

---

[97] GOOG-DOJ-04937154 at '159, "Optimized pricing in the Open Auction Comms" (Mar. 23, 2018) – Google communications document on optimized pricing in the open auction. ("AdX optimized pricing uses all event level AdX data. To price as intelligently as possible, all event level data from past Open Auctions may be considered (subject to contract terms). This means a buyer's historic bids may be used as part of the prediction model to set a price for them. Please note inputs to optimized pricing remain confidential to the buyer per the contract.")

[98] GOOG-NE-06879156 at '159, "1p Auction - Bid Data Transfer Roll-out plan" (July 24, 2019) - Internal Google presentation discussing rollout of data redactions in Bid data transfer files and impact on publishers [redacted])

Expert Report of P. Pathak, ¶ 150 ("In addition, Google broke DFP publishers' ability to measure the performance of Header Bidding. Google provides two types of data files to DFP publishers: Data Transfer files, which include Header Bidding bids, and a Bid Data Transfer file, which includes AdX and Exchange bidding bids, commonly referred to as "DT" files. Google broke the link that allowed publishers to compare the results of these files. Externally, Google described this breakage as about protecting user privacy. However, internally, Google described the data redactions as [redacted].")

[99] GOOG-NE-06879156 at '159, '167, "1p Auction - Bid Data Transfer Roll-out plan" (July 24, 2019) - Internal Google presentation discussing rollout of data redactions in Bid data transfer files and impact on publishers.

[100] GOOG-DOJ-27757105 at '106, "Re: Privacy Chat" (September 11, 2019) - Internal Google email communication discussing publisher response to Bid DT redactions.

[101] GOOG-NE-06879156 at '167, "1p Auction - Bid Data Transfer Roll-out plan" (July 24, 2019) - Internal Google presentation discussing rollout of data redactions in Bid data transfer files and impact on publishers."; GOOG-DOJ-27757105 at '106, "Re: Privacy Chat" (September 11, 2019) - Internal Google email communication discussing publisher response to Bid DT redactions. ("We had a call with [redacted] yesterday, where [redacted])

137. Google's own access to and reliance on the data advantage discussed above confirms that ML models for auction optimizations require substantial, fresh data. Google's experts are incorrect that buyers and sellers could simply use data analytics or ML to optimize their auction strategy when buyers and sellers would not have had access to the data sources required to make such data analytics or ML models effective.

138. Prof. Milgrom's report provides an ineffective solution for sellers to overcome the advantage Google Ads has. He suggests that sellers could raise AdX floors based on expected informational advantages AdX bidders have.[102] Because sellers do not have enough data, their adjustments would be mostly in the dark, especially given Google's auction manipulations such as Bernanke, DRS, RPO, EDA and throttling, which, as I discuss in more detail in the following Section V, alter bids away from their valuations.

139. In addition, Prof. Milgrom opines that sellers can improve optimizations on floor prices by working with third parties such as [redacted] or performing experiments themselves.[103] However, these solutions do not address the fact that neither sellers nor third parties have access to the fresh and relevant data required to effectively perform such experiments. Sellers who contract experiments out to third parties or conduct experiments themselves are still unable to overcome the fact that Google has a different dataset regarding bids and users for all Google auctions, while sellers still only have anonymized data from auctions they participate in, without user or bid data.

E. Impact to Buyer and Seller models

140. As I described above, buyer and seller models need access to fresh, relevant data. If the ML models had access to the same data, I would expect bids and floor prices to be similar,

---

[102] He states that "[e]ven in cases where publishers may be unable to incorporate every bit of live information to update the static prices another bidder might offer, publishers could still raise the AdX floor price to offset any *expected* informational advantage that a bidder on AdX might have." (emphasis in original) Expert Report of P. Milgrom, ¶315.

[103] Expert Report of P. Milgrom, ¶32(c) ("There are supply-side intermediaries who assist in this process, as well, including [redacted], which offers floor price optimization tools and other services for publishers to 'guarantee maximum revenues."). In the same paragraph, he notes that [redacted]

Clickio. "Maximize Revenue, Minimize Hassle, Accessed on August 28, 2024. https://clickio.com/monetization/

because the models are responding to the same variables. However, in this case, Google has more fresh, relevant data than buyers and sellers, which it uses for its ML models.

141. Google's auction manipulations then exploit the fact that buyers' and sellers' models cannot estimate as well as Google's models do. In addition, its auction manipulations increase noise in the auction results that buyer and seller models use for future estimations. The auction manipulations also increase the number of degrees of freedom for learning how to respond to changes in the auction.[104]

V. GOOGLE'S AUCTION MANIPULATIONS FURTHER LIMIT THE AMOUNT OF USEFUL INFORMATION THAT BUYERS AND SELLERS HAVE FOR MACHINE LEARNING

142. Google's experts Profs. Milgrom, Ghose, and Wiggins all opine that sellers and buyers can experiment to detect and respond to Google's auction manipulations.[105] However, this is incorrect. Google's auction manipulations (including Bernanke, DRS, RPO, and EDA), its failure to disclose some of them, and its throttling of these programs, as well as large-scale experimentation, make it impossible for sellers and buyers to optimize very well. As I will discuss below, this is because Google Ads' bids into AdX have *high bid variance* due to Google's programs. Higher variance (i.e., high noise, high uncertainty) makes estimation harder in almost any type of estimation problem.[106] Google internal documents suggest that advertisers do have difficulty coping with performance variance.[107] Google's auction manipulations can adjust bids away from buyer valuations, possibly adjusting them higher than any bidder's valuation, and often changing the winner of the auction.

---

[104] See Section VI.A for more detailed information.

[105] Expert Report of P. Milgrom, ¶32 ("And even for those programs for which details are not disclosed at all, advertisers' and publishers' routine data analysis and experimentation with bids and floor prices are typically sufficient for them to identify optimal strategies.")

Expert Report of A. Ghose, ¶158 ("In addition to relying on their ad tech providers' floor price optimizations, publishers can also test the outcomes of different floors to determine how best to optimize their revenue.")

Expert Report of N. Wiggins, ¶86 ("When one publisher was losing auctions due to high reserve prices, Google reached out proactively to suggest that the publisher use Google's automated tool for optimizing reserve prices or that it test alternative reserve prices itself using Google's manual experiments tool.")

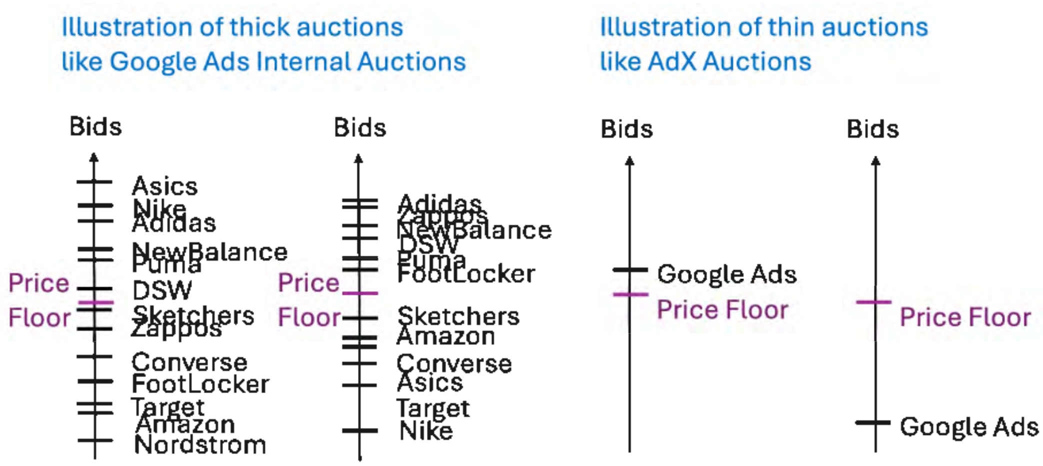
[106] Hastie, T., Tibshirani, R., and Friedman, J "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" Springer 2nd Edition, p.38 Figure 2.11. Prediction error increases with increase in variance and model complexity

[107] GOOG-DOJ-15727758 at '764, (February 10, 2014) - Internal Google meeting notes ("New idea - where does GDN variability come from? • Advertisers have a difficult time coping with performance variance. • How much of this is our experiments ("[redacted]% of traffic is going thru at least 1expt")? How much is due to other macro factors, etc? • Are these avoidable variations or unavoidable ones? If the latter, maybe we should build products.")

143. Often, Google Ads is the only bidding tool submitting bids into an AdX auction. According to an internal document, Google second prices itself in [redacted] of the winning auctions.[108] In these cases, the seller price floors need to respond only to what Google Ads might bid. However, since Google Ads' bids are high variance around the true valuations, the optimization of floor prices is difficult for sellers.

144. In Figure 7 below, I illustrate two types of auctions. One is a thick auction (left), with many similar bids at the top. The other is a thin auction (right), like an AdX auction in which only Google Ads has bid. The auction on the left might be like an internal Google Ads auction with many bidders having similar bids if they use the same Google machine learning model for valuations and bid close to their valuations. Here, the price floor does not matter since there are many bids above the floor. The seller does not see the auction on the left, they only see the auction on the right, where the choice of price floor determines the auction outcome.

Figure 7: Representation of "thick" and "thin" auctions, second price auctions. [109]



145. Figure 8 below is a Google internal document illustrating the challenge of choosing the optimal bid against another bidder that has high variance bids (like Google Ads). Prediction

[108] GOOG-AT-MDL-001386659 at '666, "gTrade Overview: GDN Leads Q2" (June 27, 2013) - Internal Presentation on gTrade, ("GDN second prices itself on [redacted] of winning queries.")

[109] Expert Report of M. Weinberg, ¶36 ("Generally speaking, thick markets are markets where the number of competitive buyers is relatively large.")

of the optimal bid is just as challenging as prediction of the optimal floor, since the floor should be set just below the highest bid; extra variance makes both problems harder.

Figure 8: Importance of Distribution for Determining Optimal Bid[110]

[redacted]

146. The extra variance arises because the bids have been changed in ways that do not reflect valuations. This makes the total information in the bid data less useful for buyers' and sellers' machine learning models.

147. While Google programs Bernanke and DRS add variance directly to the bids, Google's RPO program adds variance to the floors, which in turn leads to extra variance in the bids responding to those floors. Google programs Last Look and EDA shift the burden of estimation to the seller, who has less data than Google and cannot optimize floors very well, due in part to the high-variance bids it receives. I will discuss these programs below.

A. Bernanke

148. Below I discuss my understanding of Google's Bernanke program, which mirrors Program B discussed above in Section III.B.2. Contrary to the opinions of Google's experts, publishers

---

[110] GOOG-TEX-00841213 at '226, "AdX First Price Bidding" (October 10, 2019) - Internal Google presentation at DVA Quality Summit [redacted]

and advertisers could not create a model to optimize their strategies for the reasons discussed below and above in Section III.B.2 with respect to Program B.

1) Bernanke exploits that sellers cannot optimize floor prices

149. Google launched Project Bernanke to expand the number of auction wins in AdX by allowing Google Ads buyers to purchase otherwise unsold impressions, which made up around half of all impressions at the time Bernanke was introduced.[111] Project Bernanke creates a pool from impressions where Google Ads second-prices itself to later use on auctions where it would not ordinarily clear an auction reserve; it could also use the pool change the winner of the auction.[112] As discussed, internal documents reveal that in the second-price auction, Google Ads second priced itself [redacted] of the time.[113] The details of Bernanke were not communicated to advertisers or publishers.[114]

150. I understand that Google also was aware that large persistent "auction gaps" exist between the auction winning bid and clearing price because sellers could not set floor prices effectively.[115] In fact, in an internal communication, Google employees comment that "[t]he delta between the winning buyer's valuation and market clearing price is indeed a positive surplus. Fundamentally that surplus can be captured by the buyer or the seller. Who gets to capture depends on who has the information advantage."[116] Google engineers measured the "auction discount" in 2015 and found that [redacted]

---

[111] See GOOG-NE-13468541 at '541, "Bernanke experiment analysis" (September 3, 2013) - Internal Google document on Project Bernanke experimentation ("[redacted])

[112] GOOG-NE-06839089 at '098-099, "Project Bernanke Quantitative Easing on the Ad Exchange" (October 1, 2013) - Internal Presentation on Bernanke by gTrade Team (noting that [redacted]")

[113] GOOG-AT-MDL-001386659 at '664, "gTrade Overview: GDN Leads Q2" (June 27, 2013) - Internal Presentation on gTrade ("GDN second prices itself on [redacted] of winning queries.")

[114] Expert Report of Prof. Milgrom at ¶232 ("While the specific details of Project Bernanke were not communicated to advertisers or publishers, the strategies of bidders in auctions are routinely kept confidential for the benefit of those bidders.")

[115] Expert Report of Prof. Milgrom, fn 773. *See also* GOOG-DOJ-13203511 at '511, "Cookie based Dynamic Reserve Price Optimization (RPO) - mini PRD" (May 1, 2015) - Internal Google document on background and feature description of Cookie based Dynamic Reserve Price Optimization (RPO) ("There is often a large 'auction discount' in the AdX 2nd price auction - a gap between what buyers are bidding and what they end up paying when the auction closes.")

[116] GOOG-TEX-00777254 at '262, "Re: Op-ed that RTB should move to first-price auction -- what do you think?" (No Date), Email from [redacted] to [redacted] discussing RTB moving to first-price auction.

[redacted].[117]

151. I understand that Google had three main variants of Bernanke: "Bernanke" and "Global Bernanke," which were in the second-price AdX auction, and "The Alchemist," which was in the first-price AdX auction.[118] I also understand that Global Bernanke differed from Bernanke by maintaining a single pool across AdX publishers instead of individual per-publisher pools.[119]

152. Google referred to Bernanke as a subsidy program, stating that "advertisers on the first set of queries are net payers, paying to subsidize those net takers in the second set."[120] In addition, Google "throttled" Bernanke, meaning that it turned Bernanke off and on. Sellers and buyers never knew which queries were being subsidized, which made estimation and response by sellers and buyers more difficult, as shown in Figure 9.[121]

---

117 Expert Report of P. Milgrom, fn 773.

See also GOOG-DOJ-13203511 at '511, "Cookie based Dynamic Reserve Price Optimization (RPO) - mini PRD" (May 1, 2015) - Internal Google document on background and feature description of Cookie based Dynamic Reserve Price Optimization (RPO) ("There is often a large 'auction discount' in the AdX 2nd price auction - a gap between what buyers are bidding and what they end up paying when the auction closes. [redacted]

[118] Expert Report of M. Weinberg, ¶232 ("In this section, I analyze the conduct referred to as Project Bernanke, which was later expanded to Project Global Bernanke, and Project First Price Bernanke (also called 'The Alchemist').")

[119] Expert Report of M. Weinberg at ¶240 ("Project Global Bernanke maintained a single pool for all publishers across AdX instead of individual per-publisher pools. Project Global Bernanke aimed to keep this single pool roughly empty at the end of each billing period, implying an average take rate of 14% across all of AdX, subject to the aforementioned additional constraints such as floors on the revenue of individual publishers and on the average take rate for individual publishers.")

[120] GOOG-AT-MDL-004555192 at '194, "Auction Discussion on Incentives" (No Date) - Internal Google document containing discussion between gTrade team ([redacted]) and GDN Auction team ([redacted]) ("Advertisers on the first set of queries are net payers, paying to subsidize those the net takers in the second set.").

See also GOOG-AT-MDL-001386659 at '666, "gTrade Overview: GDN Leads Q2" (June 27, 2013) - Internal Presentation on gTrade ("Bernanke: if spending huge $$$ as subsidy, spend it wisely.")

[121] GOOG-DOJ-AT-00569945 at '947, "Auction Discussion on Incentives" (No Date) - Internal Google discussion between gTrade team ([redacted]) and GDN Auction team ([redacted]) ("Advertisers on the first set of queries are net payers, paying to subsidize those the net takers in the second set.")

Figure 9: Sellers and buyers could not identify when Bernanke ran[122]

> **(Response from [redacted] about bid lowering)**
> [redacted],
> While I agree you make a perfectly valid point (an advertiser can in theory win *THE* query where we applied DRS/Bernanke) at a lower bid, I would postulate:
>
> 1) They don't know *WHICH* random set of queries we are "subsidizing", hence lowering their bid will hurt volume on all the other queries (e.g. adsense, other adx queries where Bernanke not active). We currently limit any advertiser to being "subsidized" (margin < 14%) to at most 10% of their total queries across the whole network. So sure, they could drop bid on these 10% and possibly still win, but since they don't know which 10%, then the other 90% of queries they'd lose volume and it wouldn't be a pure win (lower price for same volume).

153. Since Google could determine when a seller's floor price will be too high or too low, Bernanke was then able to exploit this knowledge by sometimes dropping bids to the floor price and raising bids above the floor price.[123] Thus, the seller model for floor prices would never have the right information to adjust against Bernanke because it could not know how and when the bids would be adjusted.

2) Bernanke would increase variance of bids

154. Google recognized that Bernanke increased variance of bids in auction data, stating that "[redacted]"[124]

---

[122] GOOG-DOJ-AT-00569945 at '947, "Auction Discussion on Incentives" (No Date) - Internal Google discussion between gTrade team ([redacted]) and GDN Auction team ([redacted]) ("[Advertisers] don't know *WHICH* random set of queries we are 'subsidizing'…")

[123] GOOG-AT-MDL-001386659 at '666, "gTrade Overview: GDN Leads Q2" (June 27, 2013) - Internal Presentation on gTrade ("GDN second prices itself on [redacted] of winning queries." "Removing second bid would reduce GDN-> pub payout by [redacted]!")

[124] GOOG-AT-MDL-008842383 at '386, "Declaration of Nirmal Jayaram" (No Date) - Declaration of Google buy-side engineer [redacted]

155. Sellers and buyers did not receive complete information about the true second-price bid, because Bernanke changed its value before submitting the bid.[125] Bernanke increased variance in bid data by adjusting the bids either higher or lower, away from the buyer's valuation.[126] Prof. Milgrom's report does not state what effect Bernanke (or its successor program, Alchemist) had on sellers. He points out that Bernanke and Global Bernanke increased uncertainty, because these programs sometimes increased and sometimes decreased prices, either way moving them away from their valuations.[127]

156. The amount Bernanke changed bids depended on the size of the Bernanke subsidy pool and their settings for the multipliers, which used historical data and auction simulations.[128] The initial release of Bernanke had a match rate of [redacted]" That is, the first version of Bernanke successfully altered the outcome on [redacted], increasing variance.[129]

157. I understand that the variance created by Bernanke had revenue impacts that were not necessarily easy to predict in advance. [redacted]

---

[125] Expert Report of M. Weinberg, ¶233 ("The starting point for Project Bernanke is Google's observation that among auctions won by GDN advertisers, [redacted] Because GDN is 'second-pricing itself', GDN would benefit by lowering the second-highest bid it sends in order to lower the payment GDN must make to win the impression..... This aspect of Project Bernanke is akin to facilitating collusion among GDN advertisers (and in this case, without their knowledge). The second half of Project Bernanke uses the savings from the first half and spends it to subsidize overbidding.") *See also id.* at Appendix H ¶¶84-88; ¶¶94-96; ¶¶189-191; ¶¶196-199; ¶¶208-211.

[126] GOOG-TEX-00329374 at '374-375, (No Date) - Internal discussion on fee transparency ("Further the Bernanke schema is not transparent and can warp pricing and rev share on a query by query basis, which is something the industry has suspected us of doing for some time.")

GOOG-AT-MDL-008842383 at '386, "Declaration of Nirmal Jayaram" (No Date) - Declaration of Google buy-side engineer, ("[redacted])

[127] Expert Report of P. Milgrom, ¶164 ("Plaintiffs' experts overlook that the document shows [redacted] ("Note that the theoretical effects of the bid optimization programs on publisher revenues are positive or ambiguous…The theoretical effect of Bernanke and Global Bernanke on publisher revenue is ambiguous, because it lowered clearing prices for some impressions sold on AdX while increasing the prices of the extramarginal impressions it won. The theoretical effects of Alchemist on publisher revenue are also ambiguous.")

[128] Expert Report of M. Weinberg, ¶235 ("Project Bernanke computes the pair of adjustment parameters (α,β) using historical GDN data and auction simulations on a per-publisher basis in a manner that maintains an average take rate of 14% for each billing period. In order to maintain an average take-rate of 14%, Google created for each publisher a 'Bernanke pool' that is added to whenever GDN's take rate exceeds 14% and consumed from whenever GDN's take rate falls below 14%.")

[129] GOOG-DOJ-AT-00569956 at '977, "AdX + gTrade Overview" (October 1, 2014) - Internal presentation on gTrade launches, [redacted])

[redacted][133] These statistics are evidence that the uncertainty created by Bernanke changed the auction in ways that might not have been predictable in advance.

158. Google's internal documents discuss cases where bidders seemed to behave illogically when Bernanke was active, an indication that variance affected bidding models.[134] In one case, Google Ads would bid more in an Open Auction than a Private Auction when the Open Auction floor was higher than the Private Auction floor.[135] I would expect the bidding model to try to win the lower floor Private Auction, rather than the higher priced Open Auction, given it is for the same impression. Google's own engineers attributed the "weird behavior" to Bernanke.[136]

---

[130] GOOG-AT-MDL-B-002088926 at '926, "Re: Bernanke pub impact on Mobile" (December 5, 2013) - Internal Google Email between from Nirmal Jayaram to [redacted] ("I looked at pub payout and RPM changes for all pubs for mobile traffic and all traffic on 12/03/2013. [redacted]

[131]GOOG-AT-MDL-B-002514119 at '119, "Re: Pub impact list for global Bernanke" (July 29, 2015) - Internal Google email from Nirmal Jayaram to [redacted]

[132] GOOG-AT-MDL-B-002096117 at '117, "Re: Draft version of 'Pub impact estimates from CPD-based Global Bernanke" (February 13, 2015) - Internal Google email from [redacted] to [redacted]

[133] GOOG-AT-MDL-B-002096117 at '117, "Re: Draft version of 'Pub impact estimates from CPD-based Global Bernanke" (February 13, 2015) - Internal Google email from [redacted] to [redacted] ([redacted])

[134] GOOG-NE-13547436 at '437, "Re: AdX Issue" (January 17, 2017) - Internal email from [redacted] to [redacted] ("We have noticed some weird behavior on Open Auction/Deal purchase allocation for at least two buyers (Google Adwords & [redacted]). Let me try to explain: On a given scope, when one of these buyers as a PMP with a private auction floor, the more we increase the open auction floor on the same scope the less this buyer buys through this PMP (when the OA floor is superior to the PMP floor). We have been able to confirm that this weird behavior only occurs with AdWords and [redacted]. Whereas, for all the other major buyers in the same scenario the more the open auction floor increases the more they buy through the PMP.")

[135] GOOG-AT-MDL-004591613 at '626, "Advertising Revenue Playbook" (May 1, 2021) - External Google presentation on inventory monetization for publishers ("Open Auction Definition: Your ad inventory is available to all participating programmatic advertisers in a real-time bidding (RTB) online auction where the highest bidder wins the impressions;" "Private Auctions Definition: You invite an exclusive group of advertisers to bid on your inventory first through a RTB auction.")

[136] GOOG-NE-13547436 at '436, "Re: AdX Issue" (January 17, 2017) - Internal email from [redacted] to [redacted] ("As for the root cause, I suspect Bernanke might explain this. [redacted])

This "weird behavior" increased bid variance because bidders were acting differently in auctions subject to Bernanke than they had been without Bernanke.

3) Bernanke impact to buyers and sellers

159. Returning to my ML Abstracted Auction from earlier, "Program B" in the ML Abstracted Auction mirrors the auction modifications Google made for the Bernanke program. I expect Bernanke to weaken the effectiveness of machine learning for both buyers and sellers because Bernanke alters the clearing price of the auction, which is a variable that both buyers and sellers would use in machine learning models.[137]

160. As I described in my ML Abstracted Auction model, if Bernanke raises clearing prices, then the seller may respond by increasing floor prices. If the model increases floor prices, then the seller could lose revenue in the future if the floor prices are set too high, because the Bernanke subsidy may not return. Likewise, if Bernanke lowers the clearing price, the seller's model may respond by lowering floor prices in the future. If the seller lowers floor prices, the model may reduce revenue for seller. This could create a situation where no possible floor price allows the highest bidders to win.

161. For buyers using Google Ads, their buyer would not observe the changes in clearing prices in AdX, because it would still receive information only from the Google Ads auction. However, had the buyer known about the clearing prices in AdX, they might have chosen to strategically shade (i.e., lower) their bids.[138]

162. Unlike our hypothetical buyer, Bernanke could adapt to changes in the system. In a Google Chat between Nirmal Jayaram (Engineering Director at Google) and Nitish Korula (Senior

[137] Bernanke is a secret program GOOG-NE-13624783 at '785, (No Date) - Internal Google presentation on Project Bernanke GOOG-TEX-00329374 at '374-375, (No Date) - Internal discussion on fee transparency ("Further the Bernanke schema is not transparent and can warp pricing and rev share on a query by query basis, which is something the industry has suspected us of doing for some time.")

GOOG-DOJ-15445619 at '619-620, "Re: GDN Mediation Detection AM Email Followup" (June 15, 2017) - Internal email discussion between [redacted] ("With respect to Bernanke: We can't talk about this at all. I recognized this is frustrating and lacks an escalation path so I'll push this one more time, however I keep hitting a 'the first rule about Bernanke is we don't talk about Bernanke' situation".)

[138] Expert Report of M. Weinberg, ¶258 ("If advertisers knew they were participating in a non-truthful auction, they would have instead considered shading their bids. Knowing the auction format is vital information to advertisers aiming to optimize their payoff.")

Director of Engineering at Google), the two discussed how Bernanke could react in the face of changing floors and adjust bids to return to a fixed margin.[139]

B. Last Look

163. Below I discuss my understanding of Google's Last Look program, which mirrors Program LL discussed above in Section III.B.3. Contrary to the opinions of Google's experts, publishers and advertisers could not optimize their strategies for the reasons discussed below and above in Section III.B.3 with respect to Program LL.

1) Last Look shifts the burden of prediction to the publisher, who cannot estimate very well

164. In the mid-2010s, publishers began to adopt "Header Bidding" technology, which allowed publishers to call non-Google exchanges and have them compete in an auction prior to Google's auction.[140] Google did not participate in the Header Bidding auction, but Google did use the clearing price of the Header Bidding auction in its AdX auction, termed the "Last Look advantage."[141]

165. Prof. Milgrom opines that because sellers can increase the floor prices for AdX above the top header bid, there is no "advantage" for AdX and Open Bidding exchanges.[142] However,

---

[139] "GOOG-AT-MDL-003675001 at '003, (September 23, 2019) - Internal chat between Nitish Korula and Nirmal Jayaram ("It's not clear to me why the margin went down. One possibility is something changed with floors. What I meant is once this data gets into the bernanke pipeline, bernanke will lower bids and we will get to 68;" "But this just seems like the market is changing and we need to react and can do so without much negative impact.")

[140] I rely on Prof. Weinberg's characterization of Header Bidding. Expert Report of M. Weinberg, ¶94 ("Header bidding refers to a simultaneous auction technology that was developed in the early 2010s. Exchanges with real time bidding capability were present before, but header bidding enabled these exchanges to compete in an auction of auctions for the impression.")

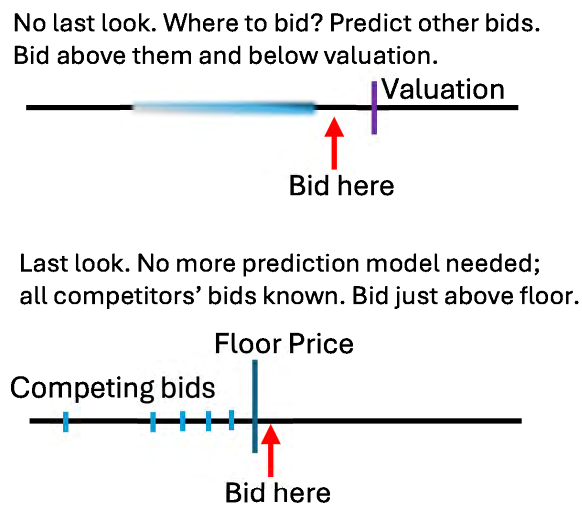
[141] Expert Report of M. Weinberg, ¶¶157-158 ("'Last Look advantage' is a phrase used to refer to AdX's ability, under Dynamic Allocation and Enhanced Dynamic Allocation, to see the header bidding clearing price before submitting their own bid. I have previously described that Dynamic Allocation and Enhanced Dynamic Allocation offer a Last Look advantage to AdX when other exchanges participate in header bidding. AdX learns the highest header bidding bid before submitting its own. This is equivalent to a first-price auction where all bidders except for AdX submit their bids first, then AdX learns the highest submitted bid and submits its own.")

[142] Expert Report of P. Milgrom, ¶516 ("As I discuss in Section X of this report, a publisher has incentives to increase ('boost' or 'inflate') the header bids it submits in Open Bidding. When it does so, it increases the probability that the header bidding bid will win the impression, which makes it harder for a bidder on AdX or an Open Bidding exchange to win the impression. Depending on the extent of this bid inflation, there can be no 'advantage' for AdX and Open Bidding exchanges. In fact, those exchanges could even be at a disadvantage.")

this is incorrect because the "Last Look" advantage shifts the burden of prediction from the bidder to the seller.

166. The illustration below shows that without a "Last Look," AdX bidders are forced to predict competitors' bids. Prof. Milgrom points out that this is an extremely difficult prediction problem.[143] With the benefit of a "Last Look," information about competitor bids and floor price are known, so no prediction is needed. If the bidder would have bid anywhere above the floor without the Last Look, they can now lower their bid to it. The price paid then depends only on the publisher's floor (which is created by the Header Bidding bid or "boosted" bid). Figure 10 below illustrates this.

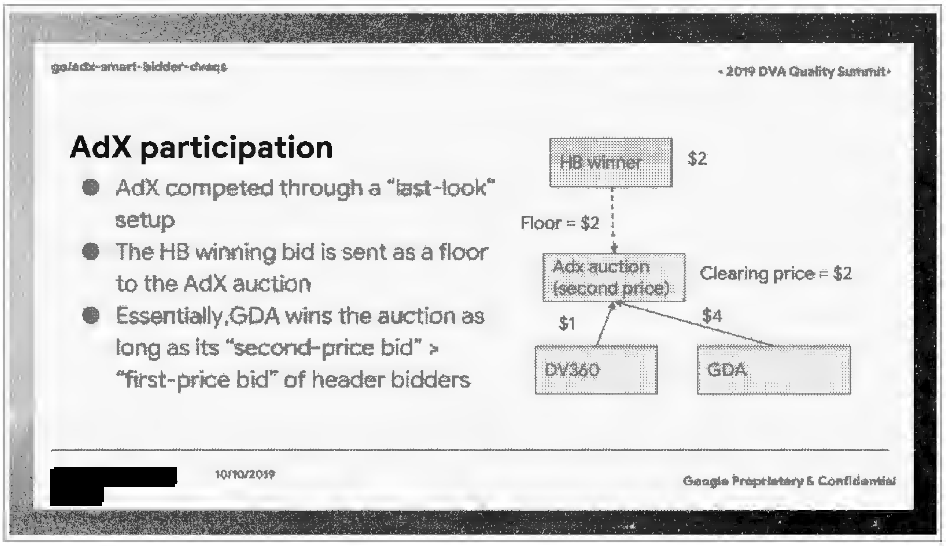
Figure 10: Last Look Advantage for Bidders



167. Last Look gives AdX buyers an advantage over sellers in prediction. AdX buyers no longer must predict others' bids, but now sellers must optimize where to place their floor. The seller can choose no prediction and give AdX the Header Bidding bid, or it can attempt to "boost" the bid by estimating a higher floor for AdX to beat, as Prof. Milgrom suggests. Google portrayed Last Look in an external document show in Figure 11.

[143] Expert Report of P. Milgrom, ¶67 ("Guessing the identities and bids of others for each different impression is a costly and error-prone activity that can lead to inefficiency when bidders' guesses are wrong.")

Figure 11: AdX Participation and Last Look[144]



168. By suggesting that sellers simply "boost bids," Prof. Milgrom ignores how estimation of floor prices is difficult for sellers as a result of Google's auction manipulations and data restrictions. Estimating where to "boost" would not be straightforward for the seller's model. If the "boost" is too high, then AdX will not clear, giving no advantage to the boost, and if the boost is too low, AdX gets a discount. Once the floor price is established, Google Ads needs to bid only $0.01 more (or choose to pass).

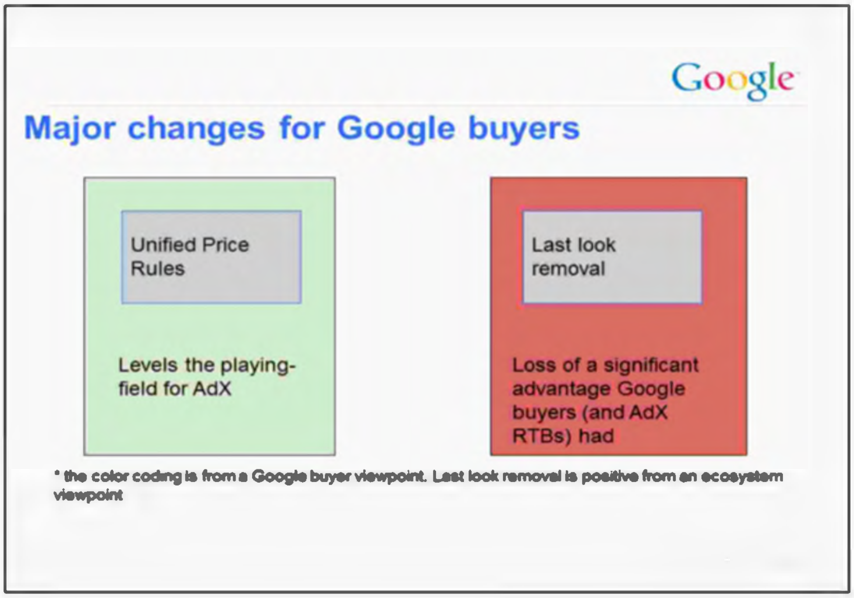
2) Last Look impact to buyers and sellers

169. As discussed, with Last Look, a buyer using AdX does not need to predict others' bids, giving an advantage over both the seller's floor price model and buyers who do not use AdX.

170. Google recognized Last Look's benefit to buyers. When Google removed Last Look as part of the move to the first-price auction in AdX, it recognized that it lost a "significant advantage," as shown in Figure 12.

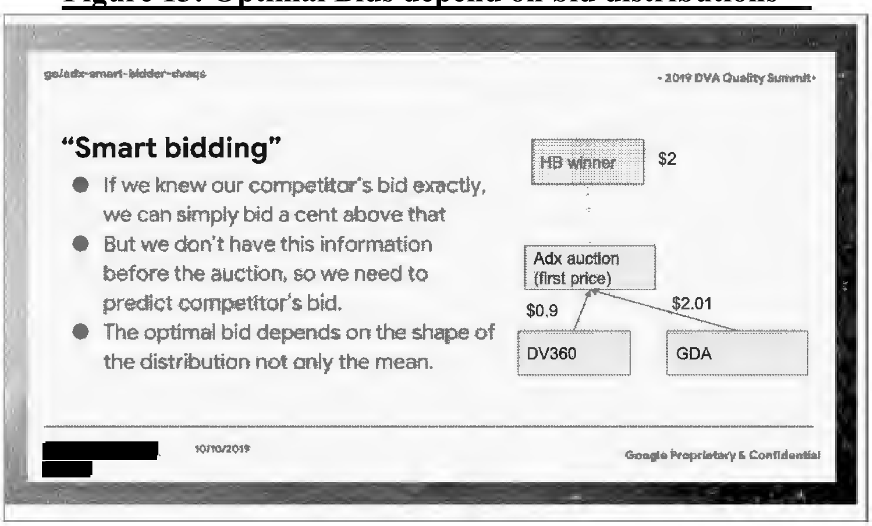
144 GOOG-AT-MDL-019771395 at '401, "AdX first-price bidding" (October 10, 2019) - Internal presentation about move to 1P Auction by Nirmal Jayaram.

Figure 12: Remove of Last Look is a Loss of Advantage[145]



171. Without Last Look, AdX would have to predict others' bids and bid above those, rather than bidding 1 cent above the floor price. In addition, Google recognized that variance in bid distributions would affect its model for optimal bids, as shown in Figure 13.

Figure 13: Optimal Bids depend on bid distributions[146]



C. Dynamic Revenue Share ("DRS")

[145] GOOG-AT-MDL-016354092 at '094, "First-price bidding Update" (September 3, 2019) – Internal presentation on losing "Last Look" and other updates.

[146] GOOG-AT-MDL-019771395 at '407, "AdX first-price bidding" (October 10, 2019) - Internal presentation about move to 1P Auction by Nirmal Jayaram.

172. Below I discuss my understanding of Google's DRS program, which mirrors Program SSD discussed above in Section III.B.4. Contrary to the opinions of Google's experts, publishers and advertisers could not create a model to optimize their strategies for the reasons discussed below and above in Section III.B.4 with respect to Program SSD.

3) DRS exploits that sellers cannot set floor prices

173. Google launched sell-side DRS to increase auction wins in AdX when bids from AdX bidders[147] would be below the floor price. Like Bernanke, in DRS, pools of money accumulate from auctions where the clearing price is much lower than the winning bid, and these are used to subsidize wins in other auctions. Sometimes the subsidies are used to permit a bid to clear a floor; other times the subsidies are used to change the winner of the auction.

174. The mechanism for DRS is an AdX adjustment of revenue share. The standard AdX revenue share is 20%.[148] On auctions where bids are high above floors, Google takes a larger revenue share (greater than 20%) to fund a pool to subsidize bids on impressions where AdX would not clear the impression, or where AdX wants to change the winner.[149] The subsidies allow AdX to lower its revenue share on these impressions.[150]

175. AdX knows it has more accurate predictions than the sellers, as Prof. Milgrom implies when he notes that AdX has "more accurate predictions than the publisher about a buyer's

[147] At the launch of DRS in 2014, AdX buyers were third-party buying tools and Google's DV360 tool.

[148] GOOG-NE-03616222 at '222, "AdX dynamic sell-side rev share (DRS v1) - project description / mini PRD" (August 1, 2014) - Internal document on DRSv1 project requirements ("Google takes a 'buy side' and 'sell side' revenue share for each AdX transaction. There is a 20% share on all transactions…")

[149] GOOG-NE-13207241 at '245, "AdX Dynamic Revshare v2: Launch Doc" (No Date) - Internal Launch Document for DRSv2, [redacted]

[150] GOOG-NE-13207241 at '245, "AdX Dynamic Revshare v2: Launch Doc" (No Date) - Internal Launch Document for DRSv2, ("The above intuition about DRS v2 is implemented by [redacted]

value"[151] and that DRS was designed to exploit AdX's (more accurate) predictions of AdX buyers' bids.[152]

176. If sellers could set floor prices optimally, DRS would not be possible, since the floor would be set just below the highest bids and the pool would not accumulate. Thus, DRS exploits that sellers cannot set floor prices optimally.

4) DRS increases variance of bids

177. Similarly, the various versions of DRS adjust the take rate of the Google AdX auction. Viewing the unevenly varying take rates as adjustments to the bids, DRS again increased variance in the bids, moving them away from the true valuations, while also introducing additional degrees of freedom.[153] Internal Google documents reveal that DRS increased uncertainty, stating that "Sell-side DRS changes the transparent, deterministic nature of the AdX second-price auction, thus adding uncertainty to buy-side simulations of bidding strategy."[154]

178. Internal Google documents show that the motivation for changing versions of DRS to tDRS (truthful DRS) was that the earlier version required bidders to "track performance across many auctions (ones on which revenue shares were discounted and others for with debt were repaid)"

[151] Expert Report of P. Milgrom, ¶419 ("An intermediary, such as AdX, with more accurate predictions than the publisher about a buyer's value can sell more impressions and increase publisher revenues (along with its own profits) by dynamically adjusting the revenue share it applies to individual impressions. Because the minimum bid required to win an impression depends on both the floor price and the revenue share applied to the impression, AdX could change the probability that an impression sold by changing its revenue share on an impression. The goal of increasing publisher revenue and enabling the sale of additional impressions is consistent with the description of the objectives for DRS in Google's internal documents: 'DRS is an optimization feature that increases seller and Google revenue by dynamically changing the AdX sell-side revenue share so that more auctions end with a winning buyer.'")

[152] Expert Report of P. Milgrom, ¶451 [redacted]

[153] Expert Report of M. Weinberg, ¶¶191, 199 ("DRSv1 impacts the auction only in one case where the highest bid is high enough to clear the reserve price, but not high enough to do so if AdX takes its full fee of 20% of the clearing price. When that happens, AdX dynamically decides to decrease its take rate so that it returns a successful bid to the ad server and wins the impression;" "Under DRSv2, AdX is allowed to either increase the take rate on a per impression basis or decrease it… AdX decides to increase or decrease the take rate based on the average take rate it applied to that publisher in the billing period. If it is close enough to the contractual requirement of 20%, then AdX sometimes decreases the take rate to win impressions that it would not have won otherwise. If it is much lower than 20% and the top bid is high enough compared to the floor, AdX increases the take rate to recoup the losses it incurred in auctions where it decreased its take rate.")

[154] GOOG-AT-MDL-008881206 at '206, gTrade Position on Sell-Side DRS (December 9, 2014) - Internal Document on Impact of DRS from GDN Perspective ("Sell-side DRS changes the transparent, deterministic nature of the AdX second-price auction, thus adding uncertainty to buy-side simulations of bidding strategy. The more aggressive the DRS, the higher the error when GDN computes the non-truthful Bernanke bids to submit for auction. Also, future GDN innovations to improve payout, matched queries, RPM, CPD, etc will become more complex, error-prone, and slower to launch.")

and perform expensive and impractical experiments.[155] Sellers would not have access to this information for any buyer and could not track which auctions had discounts and repaid debts. Thus, they would see this extra signal as variance.

5) DRS impact to buyers and sellers

179. Prof. Milgrom argues that programs such as DRS primarily allowed for sale of impressions that had floor prices that were too high.[156] However, his calculation supporting this opinion appears to be incorrect. Based on the internal document used by Prof. Milgrom to arrive at his determination that [redacted] [redacted][158]

---

[155] GOOG-NE-13205235 at '235, "Dynamic Revenue Sharing Draft" (No Date) - Internal Google document on DRS by [redacted] based on discussions with [redacted] ("…this [DRS] mechanism can make it challenging for buyers who adopt sophisticated bidding strategies which rely on learning the smallest bid with which they could have won the auction.")

Expert Report of P. Milgrom, fn. 884 ("Another source of complexity under the first two versions of DRS was that, to optimize bids, a buyer would need to track performance across many auctions (ones on which revenue shares were discounted and others for which debt were repaid), making experiments by that buyer on subsets of impressions more difficult. This was noted by Google internally as another motivation for the transition to tDRS.")

[156] Expert Report of P. Milgrom, ¶418 ("In fact, DRS applied predominantly to allow the sale of impressions where the publisher-set floor prices were too high to be met, rather than ones for which floor prices were determined by header bidding.")

[157] Expert Report of P. Milgrom, fn. 844. "DRS and RPO interaction in Simulation" (Sep. 20, 2016), GOOG-AT-MDL-007375273, at '273 [redacted])

GOOG-AT-MDL-007375273 at '273, "DRS and RPO interaction in Simulation" (No Date) - Internal Google document with statistics on DRS and RPO interaction [redacted]")

[158] Expert Report of P. Milgrom, fn. 844.

"DRS and RPO interaction in Simulation" (Sep. 20, 2016), GOOG-AT-MDL-007375273, at '273 [redacted]

Figure 14 demonstrates how DRS can change the impression winner by beating a Header Bidding bid or by beating another bid in the AdX auction. [redacted]

Figure 14: Effect of DRS on Changing Auction Outcomes[159]

[redacted]

180. Accepting Prof. Weinberg's view, if the seller knew that AdX was dynamically changing the revenue share, it likely would have responded by changing its floor price.[160] Like Bernanke, if the seller responds to DRS by increasing floor prices, then the seller could lose revenue in the future if the floor prices are set too high, because the DRS subsidy may not return (and there is no way for the seller to accurately predict if the DRS subsidy would return).

181. For buyers using AdX, DRS also affects the bidding model.[161] Prof. Weinberg explains how DRSv2 affects buyer bidding considerations: "a) if v is outside the dynamic region (i.e., v > r/0.8), the advertiser should optimally bid v, (b) if v is inside the dynamic region, shading

---

[redacted])

GOOG-AT-MDL-007375273 at '273, "DRS and RPO interaction in Simulation" (No Date) - Internal Google document with statistics on DRS and RPO interaction [redacted]

[159] GOOG-AT-MDL-007375273, at '273, "DRS and RPO interaction in Simulation" (Sep. 20, 2016).

[160] Expert Report of M. Weinberg, ¶219 ("Had they known AdX dynamically adjusted its take rate, publishers would set different price floors.")

[161] Expert Report of M. Weinberg, ¶227 ("Google misled advertisers by not revealing DRSv1, and hence led them to believe the AdX auction was a regular second-price auction, which would cause them to engage in suboptimal behavior. When advertisers believe they are participating in a regular second-price auction, they would bid their true value for the impression, because it is a truthful auction. However, DRSv1 is not truthful, as established before.")

v to some other value inside the dynamic region makes no difference, (c) if v is inside the dynamic region, the advertiser is best off omitting a bid entirely."[162]

182. As the buyers could not identify when DRS would run or when their bids would be in the dynamic zone, the buyer would not be able to make optimal adjustments in response to the program.

D. Reserve Price Optimization ("RPO")

183. Below I discuss my understanding of Google's RPO program which mirrors Program R discussed above in Section III.B.5. Contrary to the opinions of Google's experts, publishers and advertisers could not create a model to optimally respond to RPO for the reasons discussed below and above in Section III.B.5 with respect to Program R.

6) <u>RPO exploits that sellers cannot set floors</u>

184. Google launched RPO to raise floor prices on AdX buyers. Sellers could not optimize per-buyer floor prices well, as they did not have enough data. However, Google could use the combined DFP-AdX information to leverage all seller and buyer data to optimize floor prices.[163] Sellers would not have been able to get this massive amount of valuable data, since sellers compete with each other and would not want to share. For example, [redacted]'s testimony indicates that most sellers would not be willing to share their sales data with other sellers.[164]

---

[162] Expert Report of M. Weinberg, ¶226 ("DRSv2 was quite negative towards AdX advertisers' payoffs. I previously noted that from AdX advertisers' perspective, DRSv2 is a debt-aware second-price auction. In particular, a debt-aware second-price auction is exactly a second-price auction with reserve r/0.8 except if the winning bid is between r and r/0.8 it is treated as r/0.8 instead.")

[163] Deposition of [redacted] (Former Tech Lead Manager, Google), 142:6-142:6, May 23, 2024. ("Q If we can go to the next e-mail. It's from [redacted], and he writes: '[redacted]

GOOG-AT-MDL-01 9653406 at '418, "Re: [Update] [redacted] - RPO from DFP" (No Date) - Internal Google email thread ([redacted] discussing DFP and AdX merger [redacted])

[164] Deposition of [redacted] (Former Tech Lead Manager, Google), 152:3-152:7, May 23, 2024, ("THE WITNESS: Depends on the publisher. I think most publishers, if you ask them like: Hey, can we use your -- your sales information to optimize some other publisher's sales, they'd I think pretty much -- most business would say very, very quickly, no.")

185. Prof. Milgrom notes that sellers were using historical bid data to adjust floor prices, suggesting RPO was simply an automation of what sellers were already doing.[165] However, had sellers been given the same auction data as Google to be able to optimize floor prices effectively for each auction, Google's Reserve Price Optimization program (RPO) would have been unnecessary, and thus Google wouldn't have introduced RPO at all.

186. The key to RPO's success was its secrecy. RPO remained secret for one year after launch[166] (and remained obscured afterwards due to Google's throttling efforts and failure to disclose RPO's precise rules),[167] which had the effect of preventing sellers from undermining it.

7) RPO increases variance for sellers and buyers

187. RPO was a secret program for a year before it was announced on May 12th, 2016, as Prof. Weinberg discusses.[168] When Google set up RPO to automatically raise price floors that it determined were set too low, Google Ads - a major source of ad buyers - was exempt due to a secret rule AdX created that applied to Google Ads, but not to other bidders.[169] [redacted][170] This means AdX increased the price floors beyond what publishers had set for a huge number of auctions for non-Google Ads buyers, changing the distribution of winning bids that the publishers observed.

---

[165] Expert Report of P. Milgrom, ¶414 ("Even before the introduction of Optimized Pricing in April 2015, publishers were already adjusting reserve prices based on historical bids, which suggests that RPO was automating an optimization function that publishers were already doing themselves.")

[166] Prof. Milgrom opines that this generic disclosure accounts for RPO: "Ad Exchange auction model" (Aug. 24, 2014), GOOG-AT-MDL-C-000035250, at -250 ("The Google DoubleClick Ad Exchange may run limited experiments designed to optimize the auction. These experiments may include modifying the standard auction model or mechanics (e.g., a tiered, rather than second price auction); simulating ad calls and auctions; modifying the min CPM set by the publisher for an impression or otherwise adjusting publisher settings; or discounting certain bids submitted by buyers or otherwise modifying the priority of the bids submitted by buyers, in an effort to optimize the auction. Publisher's buyer/advertiser blocks will not be modified.").

GOOG-DOJ-14718372 at '372, "[ANNOUNCED] Smarter optimizations for DoubleClick Ad Exchange" (May 12, 2016) - Internal email announcement for OPA and RPO ("On Thursday afternoon we announced new optimizations for DoubleClick Ad Exchange, Optimized Private Auctions and optimized pricing in the Open Auction (aka RPO).")

[167] *See* Section VI.B ("Google's disclosures did not mitigate the data insufficiency problems that buyers and sellers faced.")

[168] Expert Report of M. Weinberg, ¶274 ("The program was launched in phases between April and October 2015. Initially, Google did not announce this program to its customers. Later, Google announced the program to its customers under the name 'optimized pricing' on May 12th, 2016, more a year after its initial rollout.")

[169] GOOG-AT-MDL-019716988 at '988, "AdX Managed Reserves" (No Date) - Internal doc on RPO ("[B]uyers that submit two bids per-auction (e.g. GDN) are already second pricing themselves and thus are exempt from RPO.")

[170] GOOG-AT-MDL-013459363 at '369, "Broadcast Media, Entertainment and Auto All Hands" (August 20, 2015) - Internal Presentation showing Mobile and Programmatic Growth, "[redacted]

188. With Google Ads exempt, floor optimization would have become much more difficult because sellers would not know that the floor price had already been changed for a large fraction of auctions, but not others.

8) RPO impact to buyers and sellers

189. RPO would make estimation more difficult for the seller, because its model would not know when RPO ran nor who RPO applied to.[171] The seller might want RPO to set all floor prices, but not realize that Google Ads bidders were exempt from the optimization.

190. RPO-eligible buyers would also have been affected. Had a buyer known about RPO and when it would have been active, it might have submitted a different bid. However, without the knowledge of RPO's existence, the buyer could not adjust its model to take into account when RPO was active.

E. Enhanced Dynamic Allocation ("EDA")

191. Below I discuss my understanding of Google's EDA program which mirrors "Program E" discussed above in Section III.B.6. Contrary to the opinions of Google's experts, publishers and advertisers could not create a model to optimize their strategies for the reasons discussed below and above in Section III.B.6 with respect to Program E.

9) EDA exploited the fact that sellers could not estimate floor prices.

192. EDA was a source of high-value impressions for Google.[172] Here, sellers and buyers had originally made a direct deal that did not involve Google. Direct deals are beneficial for sellers and buyers who do not wish to estimate the value of impressions. Each set of impressions can be viewed as a sample from a population of the seller's impressions. By purchasing a large sample, the buyer probabilistically guarantees that at least some of the impressions are high value. Importantly, neither the seller nor the buyer needs to estimate which impressions are

[171] Expert Report of M. Weinberg, ¶279 ("As one example, if Google is good at optimizing reserves via RPO, a publisher may wish to lower the reserve it sets on AdX in order to give AdX greater flexibility in optimizing its reserve, which would lead to greater revenues for both AdX and the publisher. But, via concealing RPO, Google prevented the publishers from effectively optimizing revenue.")

[172] GOOG-DOJ-15772227 at '228, "EDA Buy-Side Overview" (February 1, 2014) - Internal Google presentation overview of EDA ("Should not impact total won impressions but will give buyers access to more, valuable inventory")

high-value to estimate the value of the direct deal, because it is an average over many impressions and thus is much easier to evaluate.

193. Prof. Milgrom mentions that EDA sent only the impressions with the highest values to bidders to be sold at auction, whereas the less valuable impressions were used to satisfy the direct deal.[173] When Google removed the high-value EDA impressions and sent them to auction, it would have forced sellers and buyers to use their machine learning models to set bids and floors for these impressions, which Google knew they could not do very well.[174]

10) EDA increases variance for sellers and buyers

194. EDA placed particularly high-value impressions that otherwise would have been filled by direct deals into the auction, which increased the variance of bids the publishers expected. Publishers now needed to identify floors for these impressions, which they had not needed to do when these impressions were part of a direct deal. If the publisher did not manage to estimate the price sufficiently well, Google could place most of the impression's value in a Bernanke/DRS pool (in the case of underestimation), or the impression would go to waste (in the case of overestimation). [175]

195. Each of the programs above would contribute to the variance of the data while introducing additional degrees of freedom that make auctions on Google's platforms substantially more difficult for machine learning. This is especially true for buyers and sellers who did not have access to complete or fresh information about how Google's auctions function.

---

[173] Expert Report of P. Milgrom, ¶332 ("EDA also improved efficiency by ensuring that fewer impressions went unsold and that the impressions allocated to remnant demand were the impressions for which remnant bidders had the highest values. Because EDA assigned to guaranteed contracts the impressions with the lowest bids from remnant demand (including the ones without bids that exceed the relevant floor price, which could not be sold to remnant demand at all), EDA could often reduce the number of unsold impressions, expanding output. Similarly, the total value of impressions allocated to remnant demand increased because the impressions allocated to remnant bidders were the ones for which they had the highest bids, which are also the impressions for which they had the highest values.")

[174] GOOG-DOJ-14469279 at '279, "URGENT – Pub Removal from EDA Launch" (March 2, 2014) - Internal email on move to stop EDA for [redacted] ("I can't stress the importance of this enough. [redacted] If we turn it on there are going to be major repercussions to the relationship. Please do not proceed.")

Expert Report of M. Weinberg, fn. 114 ("[Only a] sophisticated publisher could ignore Google's suggested formulas and set the Value CPMs however they like.")

[175] Expert Report of M. Weinberg, ¶137 and fn. 181 ("In my opinion, Enhanced Dynamic Allocation likely led to an increase in win rate and increase in revenue for AdX..." and "In my opinion, the magnitude of AdX's win rate and revenue increase would be larger due to conducts such as Dynamic Revenue Sharing which increase the likelihood that AdX clears its publisher-set reserve...")

11) EDA impact to buyers and sellers

196. I understand that Google benefitted from having high-valued impressions in the auction, which serves several purposes. First, it allowed Google to use its targeting model to compel a high bid from a Google Ads buyer. This means Google would take [redacted]% of this high-valued impression's sale price.[176] It would not be able to take any of that if the impression were left to the direct deal. Second, it lowered the value of the direct deals, which reduced their values in the long run.[177]

197. Importantly, sellers could not opt out of EDA unless they wanted to opt out of AdX entirely. They were not permitted to make direct deals that avoided Google for their high-valued impressions.

F. Throttling increases variance for sellers and buyers

198. Below I discuss my understanding of Google's throttling of the programs discussed above, which mirrors the throttling discussed above in Section III.B.7. Contrary to the opinions of Google's experts, sellers and bidders could not optimize their strategies for the reasons discussed below and above in Section III.B.7.

199. Google dynamically throttles (turns on and off) its programs, further increasing variance within auctions. That is, Google programs were not evenly applied to auctions, and sellers were not notified which auctions were affected by each conduct, increasing variance and making optimization of floor prices by sellers and optimization of bids by buyers much more difficult.

200. Google's Bernanke program is one example of a program that would throttle its behavior. [redacted][178] In other words, [redacted] a buyer's

---

[176] GOOG-DOJ-15772227 at '228, 230, "EDA Buy-Side Overview" (February 1, 2014) - Internal Google presentation overview of EDA, ("Should not impact total won impressions but will give buyers access to more, valuable inventory" and [redacted].)

GOOG-AT-MDL-003407107 at '107, ("AWBID Dynamic Revshare (DRS)" (June 1, 2015) - Internal gTrade discussion on Options for AWBID DRS, "As background [redacted])

[177] [redacted]GOOG_0000120 at '121, "Enhanced Dynamic Allocation" (February 10, 2014) – External email from [redacted]

[178] GOOG-NE-12949161 at '166-167, "Bidding Strategies & Auction Mechanisms" (February 5, 2014) - Internal Presentation on Bidding Programs and their Impacts on Auctions [redacted]

bids were affected by Bernanke, without the seller or buyer knowing which bids were impacted. The uncertainty of which auctions were affected by Bernanke increased the variance for sellers and buyers.

201. Google's DRS program is another example of a program that throttled its functionality. As one example, a Google internal presentation shows DRS was launched in January 2013 and had probabilistic throttling [redacted][179] Other internal documents reveal throttling was introduced [redacted].[180] Further, as detailed by a Google internal launch document, "[redacted]"[181] A Google internal presentation on DRS confirms this behavior, stating "[redacted]."[182] In other words, if buyers and sellers deviated from their prior behavior, DRS would prevent those auctions from transacting. DRS throttling would thus increase variance for sellers and buyers.

202. Internal documents reveal that throttling, i.e., selectively changing the floor price, was an essential part of the motivation for RPO and its secrecy. Google employees remarked on this: "Let's say there's appetite for experimenting with selectively obscuring the floor price we send to buyers, and potentially doing things that don't quite fall under the description of 'truthful second price auction.'"[183]

---

[179] GOOG-AT-MDL-001386659 at '662, "gTrade Overview: GDN Leads Q2" (June 27, 2013) - Internal Presentation on gTrade , ("Dynamic Revshare (Launched 2013 [redacted]-Probabilistic throttling [redacted]")

[180] GOOG-NE-06864639 at '646, "Dynamic Sell-side Revshare on AdX" (May 9, 2014) - Internal Google document on DRS on AdX by [redacted]

[181] GOOG-NE-13199159 at '160, "Email approval from VPs for AdX DRS" (No Date) - Internal email with comments for approval of AdX DRS [redacted]

[182] GOOG-NE-13202025 at '036, "Dynamic Sell-side Revshare" (July 7, 1905) - Internal presentation on DRS by [redacted]

[183] GOOG-DOJ-32277385 at '395 On Fri, Jun 20, 2014 at 10:11 AM, [redacted] wrote: Let's say there's appetite for experimenting with selectively obscuring the reserve price we send to buyers, and potentially doing things that don't quite fall under the description of 'truthful second price auction'. We should hash out what games we want to play, write it up in a doc, and run it by [redacted].

203. At least part of Google's motivation to throttle its programs was to ensure buyers and sellers could not detect the changes its programs were producing; otherwise, for example, buyers could have reason to lower their bids.[184] As one example, a Google internal discussion states that Google can track the Bernanke performance of advertisers "to see if they notice"; one participant even adds mockingly, "Bernanke is…definitely different than 'let's do something sneaky to make more money and throttle it so nobody notices' :-)".[185]

204. Another Google internal presentation on auction optimizations gives examples for why Google needed to communicate its policies, as "Doubleclick Ad Exchange has been positioned as 'transparent' until now" and "buyer scrutiny may lead to external discovery and uncontrolled press coverage," which leads to another bullet that states "Legal risks with DRS v2."[186] In other words, auction-changing programs such as DRS were kept secret for years,[187] Google knew there were risks to not informing buyers and sellers of auction changing programs, and Google continued to keep the programs secret.

VI. BUYERS AND SELLERS COULD NOT EXPERIMENT EFFECTIVELY TO DETECT AND MITIGATE PROGRAMS

[184] GOOG-NE-12949161 at '166, "Bidding Strategies & Auction Mechanisms" (February 5, 2014) - Internal Presentation on Bidding Programs and their Impacts on Auctions ("Bernanke Throttling -Control % of times advertiser pays first price - Advertiser will otherwise have reasons to lower their maxCPC bid".)

[185] GOOG-DOJ-AT-02218556 at '559, "Untitled" (No Date) - Internal Google document containing discussion between gTrade team and GDN Auction team ("In any case, to quantitatively address these issues, we plan to holdback a set of advertisers where we won't perform any DRS/Bernanke first bid increases .... We can track the performance of such advertisers to see if they notice and/or respond to any incentive problems.")

GOOG-AT-MDL-004555192 at '195: "A: [redacted] Bernanke is totally different than randomly charging the advertiser their first price on x% of queries that they were already winning. And definitely different than "let's do something sneaky to make more money and throttle it so nobody notices" :-) "

[186] GOOG-NE-06842715 at '719, "AdX Auction Optimizations" (May 10, 2016) - Internal Presentation on AdX Auction Optimizations by [redacted] ("Doubleclick Ad Exchange has been positioned as "transparent" until now", "buyer scrutiny may lead to external discovery and uncontrolled press coverage", and "Legal risks with DRS v2".)

[187] The first iteration of DRS (DRSv1) was secret; *see* Expert Report of M. Weinberg ¶189 ("DRS was launched in August without announcing it to publishers or advertisers.")

The second version of DRS (DRSv2) was disclosed, but with caveats. See Expert Report of J. Gans ¶779 ("Google chose not to make publishers aware of the take rates for each transaction. This was an intentional choice by Google. Instead, in later implementations, publishers were given an opportunity to opt out of DRS but without proper information to make that choice").

See also Expert Report of M. Weinberg ¶197 ("DRSv2 was launched in the second half of 2016. Google announced DRSv2 when it was launched. The publishers were allowed to opt out of DRSv2, however, if they did, Google turned off DRSv1 for these publishers as well. Advertisers and ad buying tools could not opt out of DRSv2.")

205. Google's experts' opinions that buyers and sellers could experiment to detect and respond to Google programs are incorrect.[188] These arguments ignore the fact that for buyers and sellers to effectively leverage ML, they would need to have relevant fresh data at scale available for algorithms to uncover underlying patterns that occur in the auction setting. As I discussed in Section III above using my ML Abstracted Auction, buyers and sellers would not have had access to sufficient, usable data to implement such procedures. Moreover, this need for data is even more substantial if the auction is constantly modifying its procedures.

206. Buyers and sellers cannot optimize floor prices well because of the learning curve to Google's oft changing programs and their unknown degrees of freedom.

207. Google changed its conduct often, making it very challenging for sellers to respond to the fluctuating auction environment. Even Prof. Milgrom admits that policy changes take a while for buyers and sellers to adapt to,[189] and he warns that evaluations of the programs early on may not reflect later performance.[190] Of course, the learning curve would naturally be even steeper if the behavior were continually changing, as in Google's case, discussed above. Google also adjusted its programs to counter buyer and seller strategies that undermined Google's attempts to increase revenue, further complicating the efforts by those buyers and sellers to optimize auction strategy.[191]

208. Google frequently introduced new programs that could affect auction outcomes for buyers and sellers. Each of these changes meant that buyers and sellers would need to recalibrate their

---

[188] Expert Report of P. Milgrom, ¶32 ("[P]ublishers' routine data analysis and experimentation with bids and floor prices are typically sufficient for them to identify optimal strategies.")

Expert Report of S. Wiggins, ¶¶ 40-41 ("Plaintiffs' theory further fails to take into account the evidence that advertisers and publishers learn through experimentation and rely on numerous specialized intermediaries to help them implement optimal strategies... In Section II.B.1, I discuss the economic literature on learning-by-doing.")

[189] Expert Report of P. Milgrom, ¶29 ("Empirical evidence from online display advertising auctions suggests that bidders do learn to respond to auction design changes *over time, and eventually* come to adopt nearly profit-maximizing strategies.") (emphasis added)

[190] Expert Report of P. Milgrom, ¶56 ("Empirical evidence from online display advertising auctions suggests that agents learn to respond to auction design changes over time, and eventually come to adopt near-profit-maximizing strategies. This research suggests that strategic adaptation is not always immediate and that there is heterogeneity in the speed of learning, which implies that evidence about the impact of new programs gathered over short periods of experimentation must be evaluated with care: it may fail to capture eventual strategic adaptations and heterogeneity in effects over time and across agents.")

[191] GOOG-AT-MDL-B-002115457 at '548, "Re: RPO exemption mechanism change to prevent [redacted] gaming" (November 9, 2017) - Email from [redacted] to [redacted] discussing how to thwart [redacted] efforts in responding to Google's RPO program ("We discovered a little while ago that [redacted] was gaming the RPO exemption mechanism by submitting a 1 cent min-CPM, which caused them to never be reserve priced by RPO. We planned to fix this by building a more robust exemption mechanism.")

approaches to the auction to achieve their business goals—assuming they even knew about the changes. For example, notable Google program launches include the following: buy-side DRS (January 2013); Project Bernanke (November 2013); EDA (March 2014); RPO (April 2015); DRS in AdX (August 2015); a shift from second-price auctions to first-price auctions (2017); open bidding (April 2018); UPR (September 2019); and Alchemist (September 2019). In addition, Google was continually running hundreds of experiments daily, including on live traffic, which is also difficult for buyers and sellers to respond to since they do not know when the experiments occur.[192]

209. Moreover, as discussed earlier in Section V.F, many of Google's ML programs were turned on and off using throttling mechanisms, which creates variance. Throttling also made detection of Google's auction manipulations such as Bernanke and DRS more difficult.

210. Returning to my ML Abstracted Auction, a challenge for optimization based on secret programs is that the number of new variables—also known as the number degrees of freedom—that they introduce into auctions are unknown and unintuitive. Because buyers and sellers did not know which degrees of freedom would be relevant, the number of possibilities would be extremely large. For instance, buyers and sellers might wonder whether Google was performing experiments on a new program (it did that, performing live experiments on a fraction of each day's auctions);[193] whether Google was inflating some of the bids and decreasing others (it was, but the seller wouldn't know which ones);[194] whether the floors they set were not being used in the auction (which was true if RPO was on, raising the floors when Google's competitors were bidding); whether Google Ads' estimates of the valuation of the impression to each party to the auction were correct (who knows if they were, since no one else had enough data to check them); whether all of the above were true with the addition of

---

[192] Deposition of [redacted], (Software Engineer, Google), 121:24-122:6, 96:19-97:3, May 1, 2024 ("Q. Google does run hundreds of different variants of experiments, correct? A. Some features require more iteration. Yes, can be hundreds. Q. Okay, and there may be hundreds of experiments running at the same time; is that correct? A. Yes."; "Q. And I want to dig into experiments a little bit more. Those experiments that Google is conducting for yield optimization, those experiments are being done on live internet traffic and advertisements, correct? … THE WITNESS: Some experiments were run on live traffic. Some were run, we call offline, through just not live, and we did both.")

[193] Deposition of [redacted], (Software Engineer, Google), 106:6-106:12, 106:22-106:23, May 1, 2024 ("Q. Okay. And what is the range of percentages of traffic or auctions that still qualifies as—that you've seen as an experiment? … THE WITNESS: We didn't have a fixed number. But usually it's very small. It's single digit. . . . Q. So 1 to 9 percent? A. That's fair. Yeah.")

[194] I.e., Bernanke and DRS.

random throttling (turning on or performing random dramatic changes) to programs;[195] or whether Google was doing something else entirely.

211. Knowing which one (if any) of these behaviors Google engaged in mattered because buyers' and sellers' responses might be different in each case. For instance, if Google were performing random throttling, more optimizations might not help, because this type of randomness creates more variance due to noise. However, if Google were raising floor prices only for certain auctions, the seller could try to create a rule to raise up floors on the rest; provide certain buyers with discounts; or create content filters to omit certain buyers.[196] Figuring out exactly what combination of things to try would require a huge number of expensive experiments, requiring an amount of data that is exponential in the number of degrees of freedom. Prof. Milgrom's assertion that "publisher's routine data analysis and experimentation with bids and floor prices are typically sufficient for them to identify optimal strategies" is thus incorrect.[197]

212. [redacted], the former Tech Lead Manager at Google, also testified that sellers could not have optimized at the per-buyer level because they did not have enough information and could not mimic the mechanism of the auction.[198] As an auction becomes more complex, or as Google's auction manipulations reduce the amount of useful information in the auction, optimization becomes harder.

213. In high-dimensional optimization, it is typical that an optimization problem is too difficult to solve to optimality. In these cases, practitioners resort to heuristic search methods, such as

---

[195] AdX did not reduce its revenue share on each such impression; instead it "throttled" the application of DRS, meaning that it applied it probabilistically, adjusting the probability it applied DRS over time. Expert Report of P. Milgrom, fn. 857. "GOOG-DOJ-13202659, at '670 '[redacted] Presentation, "AdX DRS v1 launch review" (Feb. 13, 2015), GOOG-DOJ-13199952, at '954 '[redacted]

[196] Expert Report of P. Milgrom, ¶471 ("In reality, if a publisher wishes to exclude some types of ads, it can use content filters to achieve that result. If the publisher wishes to show the ad only when the price is sufficiently high, it can set its floor price to accomplish that too: DRS allows the ad to transact only if the publisher receives at least that floor price.")

[197] Expert Report of P. Milgrom, ¶32.

[198] Deposition of [redacted] (Former Tech Lead Manager, Google), 112:20-112:25, 113:1-113:4, 114:2-114:3, May 23, 2024 ("And if you have a single reserve price, then you could make the argument that a publisher can figure this out on their own. They could just set the reserve price to $2. Notice what fill rate they're getting. Set it to $5. Notice what fill rate they're getting. And build a graph that says: If I set it here, then here's where I seem to make more money, right? With per buyer, that's getting inside the mechanism of the auction itself. And a publisher can't do that on their own."; "And so it's something that basically only the exchange itself has enough information to really do.")

greedy methods.[199] This typically involves optimizing the most promising variables one at a time. Usually those variables are the "natural knobs" – the ones we expect to be able to adjust to improve an outcome.

214. As an example of a situation where there is an unknown number of degrees of freedom, consider the common carnival game High Striker, where a person uses a hammer to strike a lever, which hits a puck, as in Figure 15. The goal is to hit the hammer so that the puck hits a bell. An image is below:[200]

Figure 15: High Striker, Rigged Game

215. While a strong person cannot even get close to hitting the bell, a child can. Now it's your turn. What will you do? Will you try hitting the hammer harder or softer or in between? Will you hammer at different locations? Should you flip the hammer upside down? Should you hit just the corner of the hammer on the lever? All you know is that the "natural knob" to control the process—namely hitting harder on the platform—doesn't work, and that there is a huge number of possible degrees of freedom.

216. In ad auctions, raising the floor price is the natural knob that sellers can turn to remove low-quality ads. But raising the floor price may not remove low-quality ads, because Google's conduct—in this case, let's discuss RPO—has undermined the natural knob.[201] As discussed in Section V.D above, Google Ads buyers were exempt from the secret RPO program. As a

[199] Cornell University Computational Optimization Open Textbook - Optimization Wiki. "Heuristic algorithms", Accessed on August 30, 2024. https://optimization.cbe.cornell.edu/index.php?title=Heuristic_algorithms

[200] RAH Record-A-Hit Entertainment., Accessed on August 30, 2024. https://recordahit.com/wp-content/uploads/2017/07/Hi-Striker-Adult-Record-A-Hit-02.jpg

[201] RPO does not apply to all buyers; buyer networks that submit two bids are exempt from RPO, which made Google the only buyer network exempted. The two-bid exemption policy was not known to other buyer networks. "Launch Doc, 'RPO Exemption Policy V2 Launch Doc'"(Nov. 14, 2017), GOOG-DOJ-13212948, at '948 [redacted], *cited in* Expert Report of P. Milgrom, fn. 141.

result, sellers who wanted to remove low-quality ads put out by Google Ads buyers would have been unable to do so by, for example, increasing floor prices, since RPO may have set floor prices so high for non-Google Ads bidders that only Google Ads could bid above its floor anyway.

217. Similarly, Bernanke and DRS caused lower-bidding, lower-quality ads to be shown to sellers, and their natural reaction would have been to raise the reserve prices, as revealed in internal Google documents.[202] We will give an example in Section V.A.3 of when Bernanke could easily prevent the highest bidders ever from winning. An almost identical logic exists for DRS.

A. Google's disclosures did not mitigate the data insufficiency problems that buyers and sellers faced

218. Prof. Ghose opines that Google's auction manipulations – Bernanke, DRS, and RPO – were disclosed appropriately.[203] However, Google's disclosures did not provide sufficient data about program operations to allow sellers and buyers to respond by means of experimentation.

219. First, disclosures do not inform sellers and buyers about when programs are running. In fact, Google's disclosures were extremely generic. They gave little specific information to buyers and sellers, and in particular, they gave buyers and sellers no actionable knowledge about timing.[204] Without knowing if a program that adjusts bids or floors was operating, buyers

---

[202] Email from [redacted] to [redacted] "Re: Bid transparency" (Feb. 17, 2017), GOOG-AT-MDL-B-002534759, at '761, '760 ("My understanding is that you are asking about whether/when we should agree to share AdWords bids with publishers. [...] With Bernanke boost, our bids do[] not represent a true valuation and can be used to harm buyers" and "Even in a 5% sample, publishers can see AdWords high bids due to Bernanke and sur[e]ly it causes more panic on their side that they do not get their fair share. [...] So pub sees that it gets only $0.60 when AdWords valuation was $3 (Note that the real valuation was $1). They cry out louder that AdWords walked out by paying only 20% of their bid (Whereas we paid them 60%). Then they try to become smarter by putting higher reserve price for AdWords ….".)

[203] Expert Report of A. Ghose, ¶145 ("Plaintiff's experts fail to acknowledge that Google did indeed disclose its experiments and auction rules, and that any more detailed disclosures by Google or other ad tech providers would run a number of risks.")

[204] See Help Center Page, "Ad Exchange auction model" (Aug. 4, 2015), GOOG-AT-MDL-C-000035251, at -251 ("DoubleClick Ad Exchange determines the winning bidder based on the highest net bid submitted. Note that the net bid reflects any adjustments Ad Exchange may, at its discretion, have made to the bid submitted by the buyer for the purpose of optimizing the auction. [...] In some cases, the auction may close at a price lower than the reserve price applied, due to auction optimizations. Publishers are paid the Ad Exchange closing price, net of Google's revenue share, but will receive, subject to the terms governing their use of Ad Exchange, no less than the min CPM applied to the auction."). See Help Center Page, "Ad Exchange auction model" (Jun. 14, 2016), GOOG-AT-MDL-C-000035252, at '252 ("DoubleClick Ad Exchange determines the winning bidder based on the highest net bid submitted. Note that the net bid reflects any adjustments Ad Exchange may, at its discretion, have made to the bid submitted by the buyer for the purpose of optimizing the auction. [...] To optimize the auction, Google may choose to close an auction at a price lower than the reserve price that would have otherwise been applied. In such cases, the winning buyer may pay a price below the reserve and therefore receive a discount on its bid. A buyer that has received discount(s) on its bid(s) may face higher reserve prices in subsequent transactions to offset such discount(s). Subject to the terms

and sellers would have been unable to gather accurate data to feed into models regarding bids and floors.

220. Next, disclosures can increase data insufficiency problems for ML models if those disclosures are incomplete. For instance, the original disclosure of RPO did not disclose that Google Ads was exempted from RPO.[205] Prof. Milgrom's statement in his report further shows how Google's disclosures were incomplete.[206] He points out that Google had flagged in 2014 the possibility that a publisher's floor price might be modified on some impressions, but RPO was only officially announced to the public on May 12, 2016. RPO was a secret program for a year, as Prof. Weinberg discusses and as found in numerous internal documents.[207] If a seller began to rely on Google for floor optimization, that seller would have to put its trust in a black box that offered no details of how the optimization was occurring.[208]

B. Google held back the release of programs until both its buy-side and its sell-side had completed their own experiments.

---

governing their use of Ad Exchange, publishers are paid the Ad Exchange closing price, net of Google's revenue share, but will receive no less than the min CPM they specified for the auction. Unless the 'per-query revenue share' setting is enabled by a Publisher, auction optimizations may result in an auction closing at a price lower than the reserve price that would have otherwise been applied. Because the Publisher will always be paid at least its specified min CPM, the Publisher may receive more than its contracted revenue share on the transaction. In subsequent transactions, the seller's revenue share may then be reduced to offset the prior earnings in excess of the contracted revenue share, but the Publisher will always receive at least its contracted revenue share across all its Ad Exchange transactions in a given month.").

[205] Google Ad Manager. "Smarter optimizations to support a healthier programmatic market" (May 12, 2016), Accessed on September 3, 2024. https://blog.google/products/admanager/smarter-optimizations-to-suppor/

[206] Expert Report of P. Milgrom, ¶401 ("While Google had flagged as early as 2014 the possibility that a publisher's floor price might be modified on some impressions, RPO was officially announced to the public on May 12, 2016, as Optimized Pricing.")

GOOG-AT-MDL-C-000035250 at '250, "Ad Exchange auction model" (August 24, 2014) - Google disclosure of RPO ("The Google DoubleClick Ad Exchange may run limited experiments designed to optimize the auction. These experiments may include modifying the standard auction model or mechanics (e.g., a tiered, rather than second price auction); simulating ad calls and auctions; modifying the min CPM set by the publisher for an impression or otherwise adjusting publisher settings; or discounting certain bids submitted by buyers or otherwise modifying the priority of the bids submitted by buyers, in an effort to optimize the auction. Publisher's buyer/advertiser blocks will not be modified.")

[207] Deposition of [redacted] (Former Tech Lead Manager, Google), 110:16- 110:20 ("Q Going back to this e-mail from [redacted] in June 17th, 2014, we looked at the sentence where he says: "Looks like we can do RPO from DFP without pub opt-in as part of EDA provided we make some small HC changes.")

Expert Report of M. Weinberg, ¶274 ("The program was launched in phases between April and October 2015. Initially, Google did not announce this program to its customers. Later, Google announced the program to its customers under the name 'optimized pricing' on May 12th, 2016, more a year after its initial rollout.")

[208]GOOG-NE-06842715 at '721, "AdX Auction Optimizations" (May 10, 2016) - Internal Google document discussing Google conducts RPO and DRS ("Pubs complaints about coverage issues; Finger pointing at the black box.")

221. Google did not face the same challenges as others with respect to experimentation, since its buy-side and sell-side would coordinate, delaying release until both sides were ready.[209]

C. Some of Google's policies gave larger bidding tools an inherent advantage, benefiting Google's bidding tool. No amount of experimentation by competitors can remove this advantage.

222. When AdX was a 2P auction, larger bidding tools, i.e., tools that had a large proportion of advertisers, had an inherent advantage that no experimentation by others could reduce.

223. Prof. Milgrom discusses this advantage in his report,[210] observing that whenever a bidding tool hosts the highest two bids for an impression, it could drop the second bid and thus pay less than second price. That is, bidding tools were incentivized to place only one bid into each AdX auction.

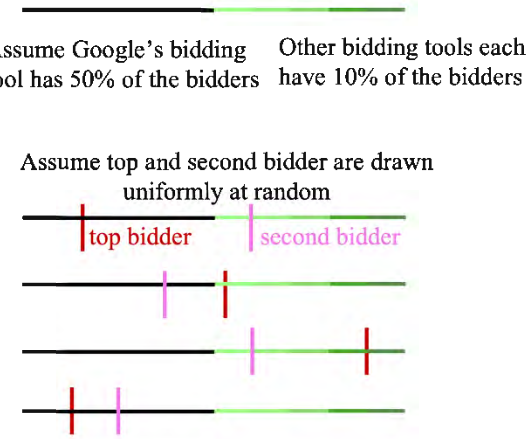
224. It is also true that larger bidding tools are more likely to have the highest two bids more often, assuming bidders are distributed among bidding tools.

225. Let us do a simulation (*see* Figure 16). Assume there are 6 bidding tools, with Google's bidding tool containing 50% of the buyers, and the 5 other bidding tools each having 10% of the buyers. We assume buyers are uniformly distributed among bidding tools (e.g., Google bidding tool wins 50% of the time since it has a 50% chance of including the winner).

[209] Deposition of [redacted] (Former Tech Lead Manager, Google), 231:1:231:8, May 23, 2024, ("Q So I understand it, do you believe that Google suppressed innovation in order to protect its existing products and markets? . . . THE WITNESS: Absolutely. Yes. I think that's clearly what happens when they say: 'Don't launch a sell-side feature until the buy side's ready, or vice versa.'"); *id* at 57:8-57:21, ("THE WITNESS: Okay. So instead of adding a competition to Google it was more kind of integrated into Google. And then they started doing things like having -- not releasing features until AdWords had already implemented support for it. And so you would tell publishers, like: Hey, we got this feature, and if you want to use it, you can wait for the other people to implement it, who they've just found out on Tuesday that we did this thing, but you could right now use AdWords, which already has it all built.")

[210] Expert Report of P. Milgrom ¶76 ("First, a DSP bidding on behalf of multiple advertisers could increase its profits by submitting only one bid into a second-price auction, instead of submitting bids on behalf of all its advertisers, while continuing to charge advertisers the same threshold prices (preserving the bidder-truthfulness of the auction). If a DSP pursues that strategy, then whenever it hosts the two highest bids for an impression, it ends up paying less for the impression than it would if it submitted bids on behalf of all its advertisers.")

Figure 16: Simulation showing that large bidding tools contain the second bidder more often.



226. In a 2P auction, when the top and second bidder come from the same bidding tool, the winner pays less than second price. Let's calculate how often this happens.

i. 50% of the time Google Ads wins, they pay the second price bid. (This is because 50% of the time, Google contains the second bidder, and the other 50% of the time they don't.)

ii. 90% of the time another bidding tool wins, it pays the second price. (There is only a 10% chance for the second bidder to land in the same bidding tool.)

227. Thus, with these simple assumptions, Google pays less than second price for half of its winning auctions just by being big. According to internal documents, Google paid the second price [redacted] of the time.[211] These documents also stated that dropping the second bid from Google's bidding tool would reduce its payout to publishers by [redacted].

228. Smaller competitors cannot use experimentation to reduce this advantage that larger bidding tools have in a 2P auction.

VII. GOOGLE POLICIES INTERFERE WITH SELLERS' AND BUYERS' LONG-TERM OPTIMIZATION GOALS

A. Google's auction manipulations interfered with sellers' ability to optimize reserve prices for long-term goals.

229. In addition to the impacts of Google's auction manipulations on the ability of sellers to optimize revenue using reserve prices, Google's auction manipulations also interfere with

[211] GOOG-AT-MDL-001386659 at '666, ("gTrade Overview: GDN Leads Q2" (June 27, 2013) - Internal Presentation on gTrade, ("GDN second prices itself on [redacted] of winning queries.")

sellers' ability to optimize for long-term goals. Google's auction manipulations optimize for auction wins, particularly for Google Ads buyers, which is not necessarily every seller's goal. For instance, let's presume a seller cared not only about getting the highest overall ad revenue, but also maximizing revenue while preserving the quality of ads on their pages. Let us also assume that sellers would like to use floor price as their mechanism for preserving ad quality.[212] For example, the *New York Times* might not want Google Ads to subsidize a low-quality ad for fake designer watches to be shown on its front page; in that case, the *New York Times* would rather leave the space empty to protect its reputation.[213] Indeed, a floor price is set to assign a default value to an ad space, a price at which a publisher would rather leave the space empty than sell it. To encourage that outcome, a seller might view the floor price as the natural knob for controlling ad quality and do so by setting a high floor price to ensure that low-quality ads, which typically bid lower, would not be chosen. Further, the seller might rely on the seller's ML model discussed above in Section III.A.2 to identify the ideal floor price to accomplish the seller's long-term quality goals.[214]

230. However, Google's subsidy programs like Bernanke (and analogously, DRS) could undermine the use of the seller ML model for this purpose because low-quality Google buyers who bid below the floor might still win by being subsidized by Bernanke or DRS.[215]

---

[212] I understand Google allows users to block ads based on certain categories and filters (Google Ad Manager Help. "Block general categories", Accessed on September 9, 2024. https://support.google.com/admanager/answer/2913554?hl=en%23:~:text=Under%2520%2522Blocks%252C%2522%2520click%2520Add,the%2520category%2520affects%2520your%2520revenue However, a seller may wish to optimize based on behavior and analytics rather than strict categories and filters created and controlled by Google. Further, a seller may be willing to compromise on ad quality at the right price in a way that categorical filters do not allow.

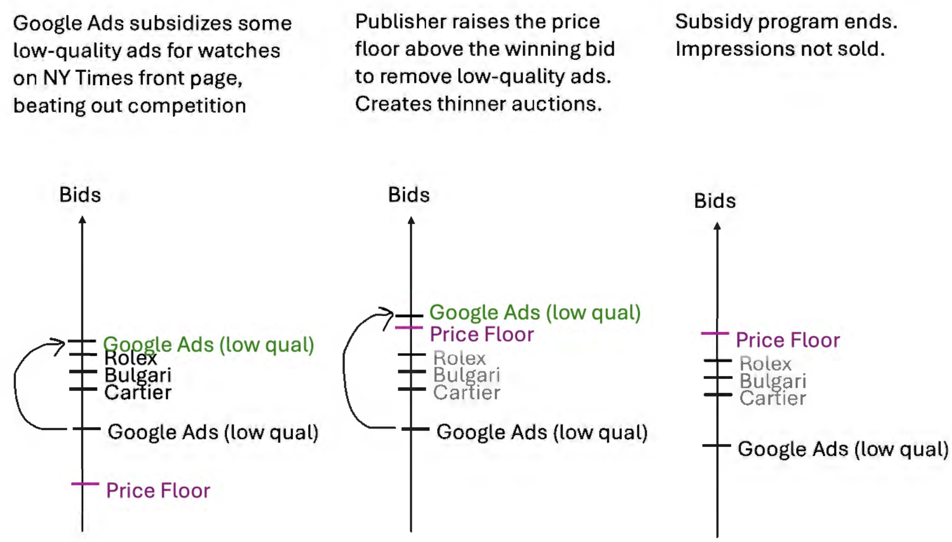
[213] See Deposition of [redacted] (Former Tech Lead Manager, Google), 99:8-99:11, May 23, 2024 (on the purpose of a reserve price) ("A. Yeah, so a publisher has inventory, they send it to the exchange, saying let people bid on this inventory, but if nobody bids above this certain price, then I would rather have no ad at all.")

[214] For clarity, I am not suggesting that ML is the only means by which sellers have or might try to optimize for ad quality. Indeed, I understand that sellers were able to optimize for ad quality using reserve prices prior to UPR. For example, Plaintiff's expert Professor Chandler explains that sellers used "personalized reserves [to] allow the publishers to screen for ad quality, by either charging more to compensate for the low quality or blocking the low quality ads altogether. UPR disables publishers from effectively using reserve prices to screen for ad quality." Expert Report of J. Chandler, ¶¶167-169. I note that UPR might further hinder the seller ML Model from identifying effective reserve prices to moderate ad quality.

[215] Expert Report of P. Pathak, ¶170 ("As I previously discussed, publishers' payoffs are affected by their users' experiences on their websites. Price floors are a tool publishers use to ensure quality advertisements on their web pages. High bids typically indicate high-quality advertisements; however, other Google programs, such as Bernanke, can subsidize low-quality advertisements, leading to transactions publishers do not want.")

231. The seller's floor optimization model, seeing low-quality ads winning impressions, might raise floor prices to omit those ads, but when it did so, and the subsidy ended, the impression would go unsold. See Figure 17 for an illustration.

Figure 17: How low-quality bids can win but high-quality bids may not.



232. In this situation, sellers' ability to set floor prices to ensure high-quality bids would be impaired *because there is no setting of the floor price that allows the highest bidders to win*. Importantly, the seller would not be aware of the timing of Google's subsidies (or even their existence), including because the seller would not have had the requisite training data as explained above in Section IV. As a result, the seller would not be able to predict when or what bids had been given subsidies, and also would not be able to predict which bids were subsidized or which bids would be subsidized in the future, as discussed earlier.

233. Prof. Milgrom describes the benefit of Project Bernanke as "positive or ambiguous" on seller revenue and dismisses internal Google reports that Bernanke led to half of sellers seeing reduced revenues.[216] He does not calculate potential long-term loss due to reputational issues.

---

[216] Expert Report of P. Milgrom, ¶157(b) ("[redacted]" *See also* GOOG-NE-13468541, at '546, "Bernanke experiment analysis" (Sep. 3, 2013) – Internal Google document sharing Bernanke experiment results, *cited in* Expert Report of P. Milgrom, fn. 289.

Expert Report of P. Milgrom, fn. 283 ("Note that the theoretical effects of the bid optimization programs on publisher revenues are positive or ambiguous. Buy-side DRS led to higher bids from Google Ads (without decreasing any other bids), which could only benefit publishers. The theoretical effect of Bernanke and Global Bernanke on publisher revenue is ambiguous, because it lowered clearing prices for some impressions sold on AdX while increasing the prices of the extramarginal impressions it won.")

234. As discussed, RPO could lead to a situation where the highest bidder did not win. If RPO created a per-buyer floor for the highest bidder that is above its bid, that bidder could not win.

235. Thus, sellers using ML models to optimize floor prices to obtain high quality ads would be unable to do so with Google's hidden programs directly subverting their optimizations.

B. Google prevented sellers from accessing data that would have reduced the burden of estimation.

236. Google's responses to seller optimizations would have reduced the effectiveness of the seller's ML models and would have diminished the seller's ability to optimize floor prices. These efforts are covered in detail in other reports I have cited below.

237. For example, Prof. Milgrom argues that in Google's view, each seller attempt to get more relevant data for optimizing floors was "harmful."[217] However, these actions were not harmful to sellers but rather buyers. Also, if sellers could set floors optimally, Google could no longer stockpile savings to subsidize other more-competitive auctions. Those stockpiled savings were unknown outside of Google and, as I have discussed elsewhere, the way they are collected and used would disadvantage a seller that is optimizing floor prices.

238. As another example, I understand that sellers used Header Bidding to perform a preliminary auction to obtain a lower bound on the floor price before calling their ad server. This lower bound would have reduced uncertainty, since the seller knows the floor should be at least the highest header bid. As Prof. Milgrom noted, Header Bidding was said to be "100X as much work" as other configurations for sellers.[218] Yet sellers did it anyway to have the

---

[217] Expert Report of P. Milgrom, ¶115 ("The huge number of new line items created as a result of publishers adopting header bidding led to a substantial increase in infrastructure costs for Google."); *id.* at ¶177 ("In October 2016, Google Ads launched Project Bell to protect its advertisers from a publisher tactic known as multi-calling, in which publishers requested bids from AdX multiple times for the same impression. If unaddressed, multi-calling would lower advertisers' profits and harm other platform participants."); *id.* at ¶521 ("In turn, publishers could exploit this advertiser multi-homing through a tactic known as price-fishing: by setting different floor prices for different exchanges, a publisher could increase its revenue at the expense of such advertisers. Internal documents suggest Google was concerned about the possibility of price-fishing after the transition to the UFPA, and UPR protected advertisers from such tactics.); *id.* at ¶537 ("[P]rice-fishing publishers impose an externality that harms advertisers and other publishers. By preventing publishers from engaging in that type of gamesmanship, UPR protected advertisers and publishers that were not price-fishing.")

[218] Expert Report of P. Milgrom, fn. 200 ("Another industry source compares header bidding to previous ad configurations: 'It's not just a little more work, it's probably 100X as much work to traffic for most publishers")

Ad Ops Insider. "Header Bidding Explained Step-by-Step" (June 8, 2015), Accessed on August 27, 2024. https://www.adopsinsider.com/header-bidding/header-bidding-step-by-step/)

benefit of better pricing data.[219] Google internal data reveals that at one point, [redacted] of Google sellers using Open Bidding sold impressions via Header Bidding.[220] sellers used Header Bidding to get information to set better floor prices. Internally, Google's stated response to the threat of Header Bidding was Open Bidding.[221] Google did not participate in Header Bidding and took steps to impede the use of Header Bidding by its sellers.[222]

239. I understand that sellers also tried multi-calling, where they sent the same impression multiple times to the auction with descending floors.[223] Multi-calling reduces uncertainty, since it allows the seller multiple trials to obtain a floor price, not just one trial. To prevent multi-calling, Google implemented Project Bell v2. Project Bell v2 detected sellers that were multi-calling and artificially capped the bids submitted to multi-calling sellers. In addition, the program disabled Bernanke, and Google prevented purchases on third-party exchanges when there was already a call for the same ad opportunity on AdX.[224] This left the seller without subsidies for setting the floor too high. [redacted]"[225]

[219] GOOG-AT-MDL-008236563 at '566, "PRD: Real-time YM with Header Container" (No Date) - DRX Sellside Product Requirements Document ("Publishers are fully buying into the view that HB can lead to better competition due to real time pricing versus averages." and "Benefits to Publishers: 'Real time' pricing vs daily averages previously used.")

[220] Expert Report of P. Milgrom, fn. 932 ("As of March 2023, [redacted])

[221] Expert Report of P. Pathak, ¶148 ("Google also responded to the threat of Header Bidding by launching Exchange Bidding. Google's Strategy Lead [redacted] referred to the 'holy grail' of Exchange Bidding goals as 'the impact of EB on the reduction of HB.'")

Defendant Google LLC's First Amended Responses and Objections to Texas Plaintiff's Third Set of Interrogatories, May 24, 2024, p.11. ("Open Bidding launched as "Exchange Bidding in Dynamic Allocation"")

[222] Expert Report of P. Pathak, ¶147, 149 ("Google did not participate in the Header Bidding, because Header Bidding threatened the control of inventory Google had built through Dynamic Allocation" and "Google also could use the ad server technology to impede the use of Header Bidding through line-item caps.")

[223] Expert Report of P. Milgrom, ¶181 ("Multi-calling with descending floor prices can effectively convert a second-price auction into a Dutch auction, in which the seller starts with a high asking price and gradually reduces the price until a buyer is willing to accept (by offering a high enough bid).")

[224] Expert Report of J. Hochstetler, ¶245 [redacted]

[225] Declaration of N. Jayaram, Aug. 5, 2023, GOOG-AT-MDL-008842383, at ¶18 ([redacted]")

240. Like multi-calling, sellers also tried (what Google calls) price-fishing, where they would call different exchanges with different reserve prices in order to optimize reserve prices. Prof. Milgrom opines that price-fishing would harm publishers that did not do it, but he does not provide data to support this opinion, instead stating it "would likely have made it more difficult for advertisers to bid optimally . . . that would harm other publishers."[226] However, both multi-calling and price-fishing are responses to variance in who is bidding and what they bid. One might think that bidders would bid similarly when the same impression appears in multiple auctions, but they do not; their bids are throttled by Google. Specifically, Google randomly assigns bidders to each auction, and its programs change their bids, as we have described in Section V.[227] Sellers would not have to price-fish or multi-call if bidders' bid distributions were low variance. However, because Google's programs increase bid variance, sellers respond with multi-calls.

VIII. GOOGLE'S ABILITY TO WIN MORE AUCTIONS THROUGH AUCTION MANIPULATIONS CREATED A COMPOUNDING OR SNOWBALL EFFECT ("SNOWBALL EFFECT") THAT ENHANCED GOOGLE'S DATA ADVANTAGE

A. Google's auction manipulations enable more AdX wins and make optimization more difficult for key auction participants

241. I understand that many of Google's auction manipulations discussed above had the effect of increasing the proportion of auctions cleared by AdX ("AdX win rates").[228] As I note in Section V above and as I further understand from Prof. Weinberg's and Prof. Gans' analysis, Google's specific programs had the following impacts:

[226] Expert Report of P. Milgrom, ¶544 ("Maintaining the ability to set exchange-discriminatory floor prices would likely have both made it more difficult for advertisers to bid optimally on Google's platform and led to externalities from price-fishing that would harm other publishers.")

[227] Expert Report of P. Milgrom, ¶90 ([redacted] ¶651 ("In practice, AdX used probabilistic throttling to avoid reducing its overall revenue share too much, and as a consequence, buyers and publishers would need to consider the possibility of throttling when determining their optimal bids and floor prices.")

[228] Expert Report of M. Weinberg, fn.107 ("I use the term 'win rate' to refer to the number of impressions won divided by the number of impressions made available in the open web display ads market.")

a. DRS in all forms increased AdX win rates, increasing the AdX match rate by [redacted]"[229]

b. Project Bernanke and Global Bernanke increased AdX win rates.[230] Between 2013 and 2015, Google analysis showed that Bernanke had increased matched queries for GDN by [redacted]). Between 2013 and 2015, Google analysis showed that Bernanke had increased matched queries for GDN by [redacted]).

c. EDA undermined direct deals and resulted in more AdX wins.[231] Exhibit 11 from Prof. Baye's report (*see* Figure 18 below) illustrates how EDA led to a reduction in the number of direct deals, which translated into more AdX auction wins.[232] Direct deals are shown in green in the figure.

---

[229] Expert Report of M. Weinberg, ¶12(d) ("Dynamic Revenue Sharing version 1 (DRSv1) increased AdX win rate and revenue and decreased non-AdX exchanges' win rates and revenues, compared to no DRS, ii. Dynamic Revenue Sharing version 2 (DRSv2), in comparison to both no DRS and DRSv1, decreased advertiser payoff, increased AdX win rate and revenue, decreased non-AdX exchange's win rates and revenues, and may also have decreased publisher revenue. iii. Truthful Dynamic Revenue Sharing increased AdX win rate and revenue and decreased non-AdX exchange's win rates compared to no DRS, iv. Google concealed information that is vital to advertisers and important to publishers by concealing DRSv1 from them.")

See also Expert Report of J. Gans, ¶806 ("By changing AdX take rate, Google was able to clear transactions it would have otherwise lost. By changing AdX take rate, Google was able to clear transactions it would have otherwise lost.1031 In the same document, Google shows the results of its DRS v1 [redacted]

[230] Expert Report of J Gans, ¶750 ("Experiments launched after the implementation of Bernanke confirm that [redacted]

Expert Report of M. Weinberg, ¶262-263 ("When combined with (Enhanced) Dynamic Allocation, Projects Bernanke and Global Bernanke enabled AdX to have a higher win rate, which would cause other exchanges to have a lower win rate.")

[231] Expert Report of M. Weinberg, ¶12(a) ("Google's implementation of Dynamic Allocation led to higher win rate and higher revenue for AdX as well as lower win rate and lower revenue for non-Google exchanges. Furthermore, Enhanced Dynamic Allocation led to an increase in win rate and increase in revenue for AdX and reduced the value of direct deals for advertisers. Reducing the value of direct deals for advertisers would decrease the revenue earned by publishers via direct deals.")

See also Expert Report of J. Gans, ¶630 ("As I explained above, EDA enables AdX (and only AdX) to transact impressions that would have been allocated to direct deals if it results in a higher clearing price. More specifically, AdX was given the ability to use the highest valued line item price as its reserve price, and transact the impression if it can beat this reserve price. No other exchange has this ability. AdX's privilege under EDA harms the competition in the ad exchange market in the long run.")

See also Expert Report of P. Milgrom, ¶157(b) ("Experiments conducted after the launch found [redacted])

[232] Expert Report of Michael R. Baye, Exhibit 11 ("Impressions on DFP for Display Ads (Narrow) Plus Direct Deals Viewed by U.S. Users Excluding House Ads June 2013 – March 2023")

d. RPO resulted in non-Google advertisers having to pay more to win. According to Google, RPO increased revenue from AdX buyers by [redacted].[233]

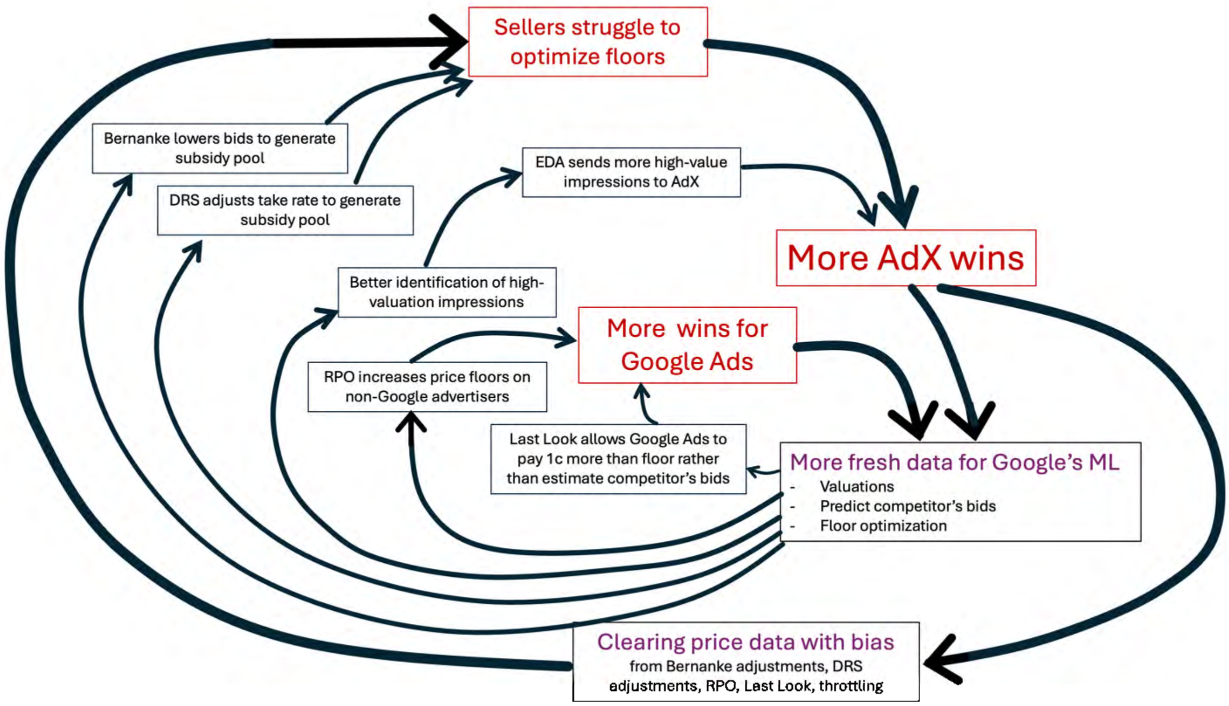
Figure 18: Exhibit 11 from Expert Report of Prof. Michael R. Baye

[redacted]

242. I have summarized my understanding of Google programs and some of their effects in Figure 19 below. I will discuss aspects of this figure in what follows

[233] GOOG-DOJ-14000011, at '012, ("AdX Dynamic Price V2" (May 26, 2015) – Internal Google presentation "Inventory RPO impact: [redacted])

Figure 19: Snowball Effect



B. When Google AdX wins more auctions, Google gathers more fresh data, which continually enables all of Google's ML programs.

243. When AdX successfully wins more auctions, it also improves the scale and freshness of the data Google collects from auction wins. Data from winning auctions is particularly valuable, since the associated data includes information on whether the user clicked on the bidder's advertisement or whether the user made a purchase (both critical inputs for Google's ML optimization models, as I discussed in Section IV.C. Thus, increased auction wins mean more fresh data, which positively reinforces Google's ML models, including models that power fine-grained targeting for buyers, predict the valuation of impressions for buyers, predicted the optimal subsidies for DRS and Bernanke, and set bids and price floors. In turn, improvements to these models would further enable Google's auction manipulations, further improve the AdX win rate, and further lead to more fresh data, reinforcing the Snowball Effect.

C. Google's auction manipulations (e.g., Bernanke and RPO) make competitors pay more to win.

244. I understand, based on my understanding of how Bernanke and RPO operate, as well as my review of Prof. Weinberg's report, that Google's auction manipulations, RPO and

Bernanke caused competitors to pay more to win.[234] For example, I understand that RPO increased price floors for non-Google buyers (but importantly, not for Google Ads buyers, who were exempted from RPO).[235] I also understand from an internal communication between the sell-side and buy-side at Google that Bernanke made GDN (Google's bidding tool) portion of AdX more competitive against third parties.[236]

245. A buyer would use data from its auction wins for future estimations, so programs that make it more expensive for non-Google buyers to win can reduce the amount of data they have available, which is likely to reduce the quality of their machine learning models in the future.

D. Throttling continues to introduce variance to the auction

246. As I discussed in Section V.F above, many of Google's auction manipulations contained throttling mechanisms that turned the programs or tools on and off at random to achieve certain goals. The randomness of the throttling behavior increased variance in auctions. Throttling thus shielded the programs from being discovered. Even if they were discovered, throttling prevented effective responses to them, as discussed. The Snowball Effect continued to enable more of Google's auction manipulations, and as Google continued to throttle these programs, throttling would continue to increase the variance in auctions, making it more difficult for buyers and sellers to optimize.

E. Google's auction manipulations and experiments continue to introduce more variance in the auctions and therefore further make optimization more difficult for auction participants.

247. As discussed above in Section VI.A, Google constantly experiments in order to develop new features and programs. It also frequently rolled out those programs, affecting auction outcomes for buyers and sellers, and throttled them to prevent their detection. With each new

[234] Expert Report of M. Weinberg, ¶253 ("Under Projects Bernanke and Global Bernanke, the win rate of non-GDN advertisers and non-GDN ad buying tools on AdX would decrease. This follows immediately because GDN submits a weakly higher bid for every AdX impression under Projects Bernanke and Global Bernanke as compared to no Bernanke. In addition, non-GDN advertisers would pay more for any impressions they still win under Projects Bernanke and Global Bernanke, for the same reason, but their total aggregate payment could still decrease due to winning fewer impressions."); *id.* at ¶281 ("RPO would lead to a payoff loss for advertisers since it leads to both a decrease in impressions won and an increase in the average price paid for impressions won. This is because the advertisers face higher reserves...")

[235] GOOG-AT-MDL-019716988 at '988, "AdX Managed Reserves" (No Date) - Internal doc on RPO ("[B]uyers that submit two bids per-auction (e.g. GDN) are already second pricing themselves and thus are exempt from RPO.")

[236] GOOG-NE-07249050 at '050, "Re: Bernanke on AdMob" (April 19, 2017) – Internal Google communication discussing effects of Bernanke. ("I believe Bernanke encourages pubs to set higher floors than they would in its absence")

program, buyers and sellers needed to recalibrate their strategies to optimize for their goals, possibly over a large number of degrees of freedom. And, as the Snowball Effect continues to enable more of Google's auction manipulations, resulting in frequent experimenting and introduction of even more programs, increasing variance in the auction, optimization continues to be difficult for buyers and sellers.

F. Google creates more and more confusing and secret programs that are difficult for everyone but Google to detect and adapt to. Google has already done experiments prior to launch.

248. As I discussed in Section VII.A, some of Google's auction manipulations work against buyer and seller goals. As I lay out in Section VI.B and V.F above, Google did not sufficiently disclose these programs and throttled them, making it difficult for buyers and sellers to detect when these programs were running. Without knowing if a program that adjusts bids or floors is operating, buyers and sellers would have been unable to harvest informative data to feed into their models for optimizing bids and floors. As the Snowball Effect continues to enable more auctions, and Google continues to insufficiently disclose these programs and throttles them, buyers and sellers will continue to struggle to detect and adapt to these programs. For each new feature or program Google rolls out, it has already conducted extensive experimentation and testing on both buy-side and sell-side.[237]

G. Buyers and sellers cannot (and will not be able to) optimize according to their own goals in the face of Google's auction manipulations

249. Many of Google's auction manipulations undermine buyers' and sellers' ability to optimize their goals. As I observe in Section VII, Google's auction manipulations interfere with the signals buyers and sellers are receiving from the auction because the programs adjust the bids and floors. Google's auction manipulations increase its own auction wins, which might not be aligned with the optimal goal of the seller or buyer. And as the snowball effect continues, and

[237] Deposition of [redacted] (Former Tech Lead Manager, Google), 57:8-57:18, May 23, 2024, ("So I understand it, do you believe that Google suppressed innovation in order to protect its existing products and markets? . . . THE WITNESS: Absolutely. Yes. I think that's clearly what happens when they say: "Don't launch a sell-side feature until the buy side's ready, or vice versa."); *id.* at 231:1-231:8 ("THE WITNESS: Okay. So instead of adding a competition to Google it was more kind of integrated into Google. And then they started doing things like having -- not releasing features until AdWords had already implemented support for it. And so you would tell publishers, like: Hey, we got this feature, and if you want to use it, you can wait for the other people to implement it, who they've just found out on Tuesday that we did this thing, but you could right now use AdWords, which already has it all built in. And so, like, a lot of the kind of -- and they would then would start delaying our feature development in order to ensure this.")

Google's auction manipulations continue to affect more auctions, buyers and sellers increasingly do not have control over their optimization goals.

H. Google's subsidies naturally create a situation where no setting of the floor price would consistently lead to highest valuation bidders winning the auction.

250. As I discuss in Sections V.A and V.C, Bernanke and DRS make it difficult for a seller to set the correct floor price. If the seller sets a floor price too low, Bernanke and DRS respond by collecting money for the pool to subsidize other auctions. Then, Bernanke and DRS allow low valuation bidders to win.

251. If a seller responds to Bernanke and DRS by raising reserve prices, Bernanke and DRS may respond by not adjusting bids to overcome the reserve price. Then, the auction ends with no winner.

252. In neither of these cases does the highest valuation bidder win.

253. Without information about when Bernanke and DRS run, the seller does not have the information to respond effectively. If the seller cannot set the correct reserve prices, Bernanke and DRS continue to create subsidy pools, and the snowball grows.

I. Buyers and sellers try to exit the snowball of Google's behavior, but Google continually blocks their exit.

254. As I discuss in Section VII.B, Google has constantly thwarted sellers' efforts to secure more information, to detect and mitigate Google's auction manipulations, and/or exit the snowball, by eliminating header bidding, punishing multi-calling and price fishing, and reducing the value of direct deals. It also throttles programs and reduces the amount of data sellers and buyers receive to conceal its operations.

J. All of the above prevent buyers and sellers from creating effective ML models.

255. As discussed above, all these different factors contribute to a Snowball Effect where Google's auction manipulations lead to more AdX wins, and more auction wins enable more fresh data which further reinforces these programs, which continue to affect auctions by increasing variance, which makes optimizations and estimations difficult for buyers and sellers. Buyers and sellers are not able to exit this snowball due to their inability to detect and adapt to these programs caused by the lack of sufficient data access, Google's incomplete

disclosures of their programs, throttling, the constant learning curve associated with the oft changing programs and Google's efforts to thwart any attempts to exit the snowball. No realistic amount of experimentation on the part of buyers or sellers can mitigate any of these factors due to the amount of variance, massive degrees of freedom, and lack of access to non-anonymized auction data.

Cynthia Rudin

Gilbert, Louis, and Edward Lehrman Distinguished Professor of Computer Science
Departments of Computer Science, Electrical and Computer Engineering,
Statistical Science, Mathematics, and Biostatistics & Bioinformatics
Duke University
[redacted] https://users.cs.duke.edu/∼cynthia/

My work focuses on **interpretable machine learning**, which is crucial for responsible and trustworthy AI. My aims are to: (1) design interpretable machine learning methods that are as accurate as black box methods, including sparse decision trees, scoring systems and interpretable neural networks, (2) apply interpretable machine learning to important societal problems in healthcare, criminal justice, energy grid reliability, and in other domains, (2) provide a theoretical foundation for why interpretable-yet-accurate models exist for most real data problems in supervised learning, (3) assist with regulation of AI, and (4) provide educational tools on interpretable machine learning (via online course materials).

I am the winner of the 2022 *Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity* from the Association for the Advancement of Artificial Intelligence (AAAI) Similar only to world-renowned recognitions, it carried a monetary reward at the million-dollar level. I am also a 2022 Guggenheim fellow, and a three-time winner of the INFORMS Innovative Applications in Analytics award based on my applied projects with domain experts. My team won second place in the prestigious 2023 Bell Labs Prize. I won the INFORMS Innovative Applications in Analytics Award three times (2013, 2016, 2019). Some accomplishments:

- The Rashomon Set Paradigm: My collaborators and I introduced a new paradigm for machine learning – instead of the algorithm providing one model (as usual), we return *all* good models, allowing the user to choose between them. This eliminates the *interaction bottleneck* to users. It allows fairness constraints and monotonicity constraints to be handled easily – by taking a simple loop over the good models. We developed this for sparse decision trees and sparse generalized additive models (GAMs). This relies on a decade of work with Margo Seltzer and our students on algorithms for optimal sparse decision trees and generalized additive models.
- Why Interpretable Models Are Accurate: Ron Parr, Lesia Semenova and I established a mathematical framework for why interpretable models are as accurate as deep learning or other black box models for most tabular data sets.
- Risk Scores: Our algorithms for sparse risk scores were applied to important healthcare and criminal justice applications: our risk score for seizure prediction in ICU patients allows for better allocation of patient monitoring resources. This work won the 2019 INFORMS Innovative Applications in Analytics Award. It helps to prevent severe brain damage in critically ill patients, and yielded the only AI model widely used in critical care brain monitoring.
- The useful PaCMAP algorithm for dimension reduction for data visualization was designed by my lab.
- I work heavily on AI regulation. I am on the National AI Advisory Committee's Subcommittee on Law Enforcement. I was on the National Academies Committee on Facial Recognition Technology that released a 2024 report.
- Energy Reliability: I led the first major effort to maintain an underground electrical distribution network using ML, in work with Con Edison in NYC. This won the 2013 INFORMS Innovative Applications in Analytics Award.
- Crime Series Analysis: My collaborators and I developed code for detecting crime series in cities. This methodology (specifically, the Series Finder algorithm) was adapted by the NYPD and their application (Patternizr) has been running live in NYC since 2016 for determining whether each new crime is related to past crimes.
- Margins for AdaBoost: I solved a well-known (previously open) theoretical problem in machine learning as a PhD student, which is whether AdaBoost maximizes the ℓ_1- margin. Subsequent work solved a COLT open problem.
- I enjoy competing in data science competitions and coaching student teams. We have won awards including the FICO Recognition Award for the Explainable Machine Learning Challenge (2018), NTIRE Superresolution Competition (2018), PoeTix Literary Turing Competition (2018), and American Statistical Association Data Challenge Expo Student Competition (2022 and 2023).
- I have given invited/keynote/plenary talks at INFORMS, KDD (2014 and 2019), AISTATS, ECML-PKDD, ML in Healthcare, FAT-ML (Fairness, Accountability, and Transparency), SIAM International Conference on Data Mining (SDM), and the Nobel Conference. My work appears in the media, including the NY Times, Washington Post, Boston Globe, Wall Street Journal, MSNBC, and National Public Radio.
- I am writing an e-book on introductory machine learning called "Intuition for the Algorithms of Machine Learning," which teaches interpretable machine learning as part of the basic ML curriculum, linked on my website.

Appendix B: Materials Relied Upon & Materials Considered

I. MATERIALS RELIED UPON

A. Expert Reports

1. 2023.12.22 Expert Report of Ramamoorthi Ravi, PhD
2. 2024.06.07 Expert Report of Joshua Gans, PhD
3. 2024.06.07 Expert Report of Matthew Weinberg, PhD
4. 2024.06.07 Expert Report of Jacob Hochstetler, PhD
5. 2024.06.07 Expert Report of John Chandler, PhD
6. 2024.06.07 Expert Report of Parag Pathak, PhD
7. 2024.07.30 Expert Report of Anindya Ghose, PhD
8. 2024.07.30 Expert Report of Paul R. Milgrom, PhD
9. 2024.07.30 Expert Report of Steven N. Wiggins, PhD
10. 2024.08.06 Expert Report of Michael R. Baye
11. 2024.08.06 Expert Report of Jason Nieh, PhD

B. Depositions and ROG Responses

1. 2024.05.01 Deposition of [redacted] and exhibits
2. 2024.05.23 Deposition of [redacted] and exhibits
3. 2024.05.24 Defendant Google LLC's First Amended Responses and Objections to Texas Plaintiff's Third Set of Interrogatories
4. 2024.08.05 Declaration of [redacted], GOOG-AT-MDL-008842383

C. Documents from Production

1. [redacted]_GOOG_0000120
2. GOOG-AT-MDL-001004706
3. GOOG-AT-MDL-001283820
4. GOOG-AT-MDL-001386659
5. GOOG-AT-MDL-001391101
6. GOOG-AT-MDL-002295716
7. GOOG-AT-MDL-003291718
8. GOOG-AT-MDL-003407107
9. GOOG-AT-MDL-003675001
10. GOOG-AT-MDL-004080757
11. GOOG-AT-MDL-004242638
12. GOOG-AT-MDL-004288674

13. GOOG-AT-MDL-004434946
14. GOOG-AT-MDL-004555192
15. GOOG-AT-MDL-004555192
16. GOOG-AT-MDL-004591613
17. GOOG-AT-MDL-006597606
18. GOOG-AT-MDL-007375273
19. GOOG-AT-MDL-007418936
20. GOOG-AT-MDL-008236563
21. GOOG-AT-MDL-008842383
22. GOOG-AT-MDL-008881206
23. GOOG-AT-MDL-009013192
24. GOOG-AT-MDL-009013263
25. GOOG-AT-MDL-009013510
26. GOOG-AT-MDL-009013567
27. GOOG-AT-MDL-012682959
28. GOOG-AT-MDL-012685117
29. GOOG-AT-MDL-012760228
30. GOOG-AT-MDL-013459363
31. GOOG-AT-MDL-016346558
32. GOOG-AT-MDL-016354092
33. GOOG-AT-MDL-016354429
34. GOOG-AT-MDL-018520983
35. GOOG-AT-MDL-018621551
36. GOOG-AT-MDL-019245873
37. GOOG-AT-MDL-019528391
38. GOOG-AT-MDL-019593106
39. GOOG-AT-MDL-019653406
40. GOOG-AT-MDL-019716988
41. GOOG-AT-MDL-019771395
42. GOOG-AT-MDL-B-001391461
43. GOOG-AT-MDL-B-002088293
44. GOOG-AT-MDL-B-002088926
45. GOOG-AT-MDL-B-002096117
46. GOOG-AT-MDL-B-002115457
47. GOOG-AT-MDL-B-002186045
48. GOOG-AT-MDL-B-002514119
49. GOOG-AT-MDL-B-002534759
50. GOOG-AT-MDL-B-002838050
51. GOOG-AT-MDL-B-003207193
52. GOOG-AT-MDL-B-005177744
53. GOOG-AT-MDL-B-005714587
54. GOOG-AT-MDL-B-005785771
55. GOOG-AT-MDL-C-000035250
56. GOOG-AT-MDL-C-000035251
57. GOOG-AT-MDL-C-000035252
58. GOOG-DOJ-03151263
59. GOOG-DOJ-04937154
60. GOOG-DOJ-13199952
61. GOOG-DOJ-13203511
62. GOOG-DOJ-13991370
63. GOOG-DOJ-14421383
64. GOOG-DOJ-14469279
65. GOOG-DOJ-14718372
66. GOOG-DOJ-15445619
67. GOOG-DOJ-15631978
68. GOOG-DOJ-15727758
69. GOOG-DOJ-15772227
70. GOOG-DOJ-27757105
71. GOOG-DOJ-28486313
72. GOOG-DOJ-28501885
73. GOOG-DOJ-31322997
74. GOOG-DOJ-32277385

75. GOOG-DOJ-AT-00060049
76. GOOG-DOJ-AT-00569945
77. GOOG-DOJ-AT-00569956
78. GOOG-DOJ-AT-01130527
79. GOOG-DOJ-AT-02190535
80. GOOG-DOJ-AT-02193461
81. GOOG-DOJ-AT-02218556
82. GOOG-DOJ-AT-02427326
83. GOOG-DOJ-AT-02493495
84. GOOG-DOJ-AT-02504615
85. GOOG-NE-02215711
86. GOOG-NE-03616222
87. GOOG-NE-04933892
88. GOOG-NE-06230051
89. GOOG-NE-06839089
90. GOOG-NE-06842715
91. GOOG-NE-06864639
92. GOOG-NE-06879156
93. GOOG-NE-07249050
94. GOOG-NE-12949161
95. GOOG-NE-13197548
96. GOOG-NE-13198969
97. GOOG-NE-13199159
98. GOOG-NE-13200831
99. GOOG-NE-13202025
100. GOOG-NE-13202877
101. GOOG-NE-13203930
102. GOOG-NE-13204103
103. GOOG-NE-13204729
104. GOOG-NE-13205235
105. GOOG-NE-13207241
106. GOOG-NE-13220268
107. GOOG-NE-13226622
108. GOOG-NE-13319630
109. GOOG-NE-13468541
110. GOOG-NE-13547436
111. GOOG-NE-13624783
112. GOOG-TEX-00329374
113. GOOG-TEX-00777254
114. GOOG-TEX-00841213
115. GOOG-TEX-00873439

D. Public Sources

1. Ad Ops Insider. "Header Bidding Explained Step-by-Step" (June 8, 2015), Accessed on August 30, 2024. https://www.adopsinsider.com/header-bidding/header-bidding-step-by-step/.

2. Authorized Buyers Help. "Bid data sharing", Accessed on July 24, 2024. https://support.google.com/authorizedbuyers/answer/2696468?hl=en

3. Authorized Buyers. "Real-time Bidding", Accessed on August 19, 2024. https://developers.google.com/authorized-buyers/rtb/openrtb-guide

4. Auxier, B., Anderson, M., "Social Media Use in 2021," Accessed September 7, 2024. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

5. Censius Blog. "How To Conduct A/B Testing In Machine Learning?", Accessed on September 4, 2024. https://censius.ai/blogs/how-to-conduct-a-b-testing-in-machine-learning

6. ClearCode. "The Main Technology Platforms and Intermediaries in the Digital Advertising Ecosystem" (2017), Accessed on June 25, 2024. https://adtechbook.clearcode.cc/adtech-platforms-and-intermediaries/

7. ClearCode. "What Is a Demand-Side Platform (DSP) and How Does It Work?" (January 31, 2024), Accessed on June 25, 2024. https://clearcode.cc/blog/demand-side-platform/

8. ClearCode. "What is an Ad Server and how does it work?" (April 20, 2024), Accessed on June 25, 2024. https://clearcode.cc/blog/what-is-an-ad-server/#toc-label-3

9. Clickio. "Maximize Revenue, Minimize Hassle", Accessed on August 30, 2024. https://clickio.com/monetization/

10. ClimateStations.com. "ny2024.gif", Accessed on August 30, 2024. https://www.climatestations.com/wp-content/uploads/2024/08/ny2024.gif

11. Cornell University Computational Optimization Open Textbook - Optimization Wiki. "Heuristic algorithms", Accessed on August 30, 2024. https://optimization.cbe.cornell.edu/index.php?title=Heuristic_algorithms

12. Dogs of the Dow. "ALL Stock Price Today (plus 7 insightful charts)", Accessed on August 30, 2024. https://www.dogsofthedow.com/stock/all.htm

13. DoubleClick Ad Exchange. "Profiting from Non-Guaranteed Advertising: The Value of Dynamic Allocation & Auction Pricing for Online Publishers" (2010), Accessed August

30, 2017. https://web.archive.org/web/20170830013310/http:/static.googleusercontent.com/media/www.google.com/en/us/adexchange/DC_Ad_Exchange_WP_100713.pdf

14. DoubleClick. “DoubleClick Advertising Exchange”, Accessed on June 25, 2024. https://web.archive.org/web/20071001100309/http://www.doubleclick.com/products/advertisingexchange/index.aspx

15. DuckDuckGo, “tracker-radar/entities/Google LLC.json,” Accessed September 7, 2024. https://github.com/duckduckgo/tracker-radar/blob/main/entities/Google%20LLC.json.

16. Encyclopedia of Mathematics. "Bernstein inequality", Accessed on August 30, 2024. https://encyclopediaofmath.org/wiki/Bernstein_inequality

17. Google Ad Manager Help. "Block general categories", Accessed on September 9, 2024. https://support.google.com/admanager/answer/2913554?hl=en%23:~:text=Under%2520%2522Blocks%252C%2522%2520click%2520Add,the%2520category%2520affects%2520your%2520revenue

18. Google Ad Manager. "Smarter optimizations to support a healthier programmatic market" (May 12, 2016), Accessed on September 3, 2024. https://blog.google/products/admanager/smarter-optimizations-to-suppor/

19. Google Ads Help. "Evaluate ad performance on the Display Network", Accessed on September 8, 2024. https://support.google.com/google-ads/answer/2404178?hl=en

20. Google for Developers. "Experiments" (July 17, 2024), Accessed on September 6, 2024. https://developers.google.com/machine-learning/managing-ml-projects/experiments

21. Google Support. "About Smart Bidding", Accessed on September 4, 2024. https://support.google.com/google-ads/answer/7065882?hl=en

22. IETF. “Cookies: HTTP State Management Mechanism”, Accessed September 9, 2024. https://www.ietf.org/archive/id/draft-ietf-httpbis-rfc6265bis-07.html#name-acknowledgements

23. Kaspersky. “What are Cookies?”, Accessed June 25, 2024. https://usa.kaspersky.com/resource-center/definitions/cookies

24. Math is Fun. "Exponential Growth and Decay", Accessed on August 30, 2024. https://www.mathsisfun.com/algebra/exponential-growth.htML

25. Nokia Bell Labs. "Bell Labs Prize", Accessed on September 6, 2024. https://www.bell-labs.com/collaboration-opportunities/prize/

26. npr. "Which Is Greater, The Number Of Sand Grains On Earth Or Stars In The Sky?" (August 30, 2024), Accessed on August 30, 2024. https://www.npr.org/sections/krulwich/2012/09/17/161096233/which-is-greater-the-number-of-sand-grains-on-earth-or-stars-in-the-sky

27. RAH Record-A-Hit Entertainment. Accessed on August 30, 2024. https://recordahit.com/wp-content/uploads/2017/07/Hi-Striker-Adult-Record-A-Hit-02.jpg

28. Scribbr. "How to Find Degrees of Freedom | Definition & Formula”, Accessed on August 30, 2024. https://www.scribbr.com/statistics/degrees-of-freedom/

29. SingHealth. "Figure 1 EEG seizure: An ictal EEG rhythm seen predominantly in the midline and left frontocentral elecrodes", Accessed on August 30, 2024. https://www.singhealth.com.sg/sites/shcommonassets/Assets/MedicalNews/MedicalNews-2019/non-compulsive-status-epilepticus-eeg-seizure-nni.jpg

30. Sky & Telescope: The Essential Guide to Astronomy. "95 Sunrises Along the Columbia Riverbank" (May 28, 2020), Accessed on August 30, 2024. https://skyandtelescope.org/observing/stargazers-corner/95-sunrises/

31. Spread Privacy. "DuckDuckGo Tracker Radar Exposes Hidden Tracking", Accessed on September 9, 2024. https://spreadprivacy.com/duckduckgo-tracker-radar/

32. statcounter, "Search Engine Market Share United States of America", Accessed September 7, 2024. https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america.

33. Statista, "Most popular mapping apps in the United States as of April 2018, by reach", Accessed September 7, 2024. https://www.statista.com/statistics/865419/most-popular-us-mapping-apps-ranked-by-reach/.

34. Statista. "Market share of leading internet browsers in the United States and worldwide as of August 2024" (June 21, 2024), Accessed on September 7, 2024. https://www.statista.com/statistics/276738/worldwide-and-us-market-share-of-leading-internet-browsers/

35. TechTarget. "Third-party cookie" (May 2023), Accessed August 30, 2024. https://www.techtarget.com/whatis/definition/third-party-cookie

36. Towards Data Science. "Curse of Dimensionality — A "Curse" to ML" (August 10, 2021), Accessed on August 30, 2024. https://towardsdatascience.com/curse-of-dimensionality-a-curse-to-machine-learning-c122ee33bfeb

37. Wikipedia. "Chebyshev's inequality", Accessed on August 30, 2024. https://en.wikipedia.org/wiki/Chebyshev's_inequality

38. Wikipedia. "Economies of Scale", Accessed on September 3, 2024. https://en.wikipedia.org/wiki/Economies_of_scale

39. Wired. "8 Google Employees Invented Modern AI. Here's the Inside Story" (March 20, 2024), Accessed on September 4, 2024. https://www.wired.com/story/eight-google-employees-invented-modern-ai-transformers-paper/

E. Books and Papers

1. Altman, N. and Krzywinski, M. "The curse(s) of dimensionality" Nature Methods.

2. Berisha, V., Krantsevich, C., Hahn, P.R., Hahn, S., Dasarathy, G., Turaga, P., Liss, J. "Digital medicine and the curse of dimensionality" npj Digital Medicine volume 4, Article number: 153 (2021).

3. Budach, Lukas, et al. "The effects of data quality on machine learning performance." arXiv preprint arXiv:2207.14529 (2022).

4. Cunningham, P., Cord, M., Delany, S.J "Supervised Learning " Machine Learning Techniques for Multimedia. Cognitive Technologies.

5. Devlin, J., Chang, M., Lee, K., & Toutanova, K. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding".

6. Domingos, P. "A Few Useful Things to Know About Machine Learning" Communications of the ACM, vol. 55, issue 10. pgs. 78-87.

7. Easley, D. and Kleinber, J. "Networks, Crowds, and Markets: Reasoning about a Highly Connected World" Cambridge University Press, 2010. "Chapter 9: Auctions"

8. Easley, D. and Kleinber, J. "Networks, Crowds, and Markets: Reasoning about a Highly Connected World" Cambridge University Press, 2010. "Chapter 9: Auctions"

9. Golestani, A., Gras, R. "Can we predict the unpredictable?" Scientific Reports 4, Article Number: 6834 (2014).

10. Gudivada, V., Apon, A., and Ding, J. "Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations" International Journal on Advances in Software, vol 10 no 1 & 2, year 2017.

11. Gupta, S., and Gupta, A. "Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review" Procedia Computer Science 161 (2019) 466–474.

12. Hagiu, Andrei, and Julian Wright. "Data-enabled learning, network effects, and competitive advantage." The RAND Journal of Economics 54.4 (2023): 638-667.

13. Hastie, T., Tibshirani, R., and Friedman, J "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" Springer 2nd Edition, p.38 Figure 2.11.

14. Kohavi, R., Tang, D., Xu, Y. "Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing " Cambridge University Press 2020.

15. Levy, P. S., & Lemeshow, S. "Sampling of populations: methods and applications" John Wiley & Sons.

16. Nicholas, J., Herbert Chan, H.W., Baker, M.A.B. "Machine learning: applications of artificial intelligence to imaging and diagnosis" Biophysical reviews, 11(1), 111–118.

17. Paninski, L. "Probability inequalities " (Columbia University Course) Statistics 4109: Probability and Statistics (Fall 2010).

18. Prapas, I., Derakhshan, B., Mahdiraji, A.R. et al. Continuous Training and Deployment of Deep Learning Models. Datenbank Spektrum 21, 203–212 (2021). https://doi.org/10.1007/s13222-021-00386-8.

19. Rudin, C. "Fundamentals of Learning Course Notes" Intuition for the Algorithms of Machine Learning: A Multimedia Textbook. https://users.cs.duke.edu/~cynthia/CourseNotes/ConceptsOfLearningNotes.pdf

20. Sarker, I.H. "Machine Learning: Algorithms, Real-World Applications and Research Directions" SN Computer Science (2021) 2:160.

21. Sarkis, A. "Training Data for Machine Learning: Human Supervision from Annotation to Data Science" O'Reilly Media, Inc.

22. Singh, Ajay S., and Micah B. Masuku. "Sampling techniques & determination of sample size in applied statistics research: An overview." *International Journal of economics, commerce and management* 2.11 (2014): 1-22.

23. UC Davis Mathematics. "Chapter 2: Curses, Blessings, and Surprises in High Dimensions", Accessed on August 30, 2024. https://www.math.ucdavis.edu/~strohmer/courses/270/Chapter2.pdf

24. van der Ploeg, T., Austin, P.C. & Steyerberg, E.W. "Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints" BMC Medical Research Methodology volume 14, Article number: 137 (2014).

25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin L. "Attention is All you Need" 31st Conference on Neural Information Processing Systems (NIPS 2017).

26. Yu, C.H. "Degrees of Freedom" International Encyclopedia of Statistical Science.

27. Zhang, W. Yuan, S., Wang, J. "Optimal Real-Time Bidding for Display Advertising" KDD '14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.

II. MATERIALS CONSIDERED

A. Discovery Responses

All available discovery responses produced within the matter of *The State of Texas, et al. v. Google*, Case Number: 4:20-cv-00957-SDJ, including:

1. The Parties' amended initial disclosures;
2. The Parties' discovery responses and objections to Interrogatories, Requests for Admission, and Requests for Production; and
3. Google's written responses to Plaintiffs' Rule 30(b)(6) Notice.

B. Pleadings

The live pleadings (complaint and answer) within the matter of *The State of Texas, et al. v. Google*, Case Number: 4:20-cv-00957-SDJ, including the Fourth Amended Complaint

C. Deposition Transcripts & Exhibits

All available deposition transcripts and exhibits within the matter of *The State of Texas, et al. v. Google*, Case Number: 4:20-cv-00957-SDJ, including:

1. Deposition and Exhibits of [redacted], April 1, 2024
2. Deposition and Exhibits of [redacted], April 3, 2024
3. Deposition and Exhibits of [redacted], April 12, 2024
4. Deposition and Exhibits of [redacted], April 17, 2024
5. Deposition and Exhibits of [redacted], April 19, 2024
6. Deposition and Exhibits of [redacted], April 23, 2024
7. Deposition and Exhibits of [redacted], April 26, 2024
8. Deposition and Exhibits of [redacted], April 26, 2024
9. Deposition and Exhibits of [redacted], April 29, 2024
10. Deposition and Exhibits of [redacted], April 30, 2024
11. Deposition and Exhibits of [redacted], May 1, 2024
12. Deposition and Exhibits of [redacted], May 1, 2024
13. Deposition and Exhibits of [redacted], May 2, 2024
14. Deposition and Exhibits of [redacted], April 5, 2024
15. Deposition and Exhibits of [redacted], May 2, 2024
16. Deposition and Exhibits of [redacted], May 10, 2024
17. Deposition and Exhibits of [redacted], May 15, 2024
18. Deposition and Exhibits of [redacted] May 17, 2024
19. Deposition and Exhibits of [redacted] Vol 1, April 26, 2024
20. Deposition and Exhibits of [redacted] Vol 2, May 21, 2024
21. Deposition and Exhibits of [redacted], May 21, 2024
22. Deposition and Exhibits of [redacted], May 22, 2024

23. Deposition and Exhibits of [redacted], May 23, 2024
24. Deposition and Exhibits of [redacted], May 24, 2024
25. Deposition and Exhibits of [redacted] Vol 1, April 19, 2024
26. Deposition and Exhibits of [redacted] Vol 2, May 2, 2024
27. Deposition and Exhibits of [redacted] Vol 3, May 3, 2024
28. Deposition and Exhibits of [redacted] Vol 4, May 24, 2024
29. Deposition and Exhibits of [redacted]
30. Deposition and Exhibits of [redacted]
31. Deposition and Exhibits of [redacted]
32. Deposition and Exhibits of [redacted]
33. Deposition and Exhibits of [redacted]
34. Deposition and Exhibits of [redacted]
35. Deposition and Exhibits of [redacted]
36. Deposition and Exhibits of [redacted]
37. Deposition and Exhibits of [redacted]
38. Deposition and Exhibits of [redacted]
39. Deposition and Exhibits of [redacted]
40. Deposition and Exhibits of [redacted]
41. Deposition and Exhibits of [redacted]
42. Deposition and Exhibits of [redacted]
43. Deposition and Exhibits of [redacted]
44. Deposition and Exhibits of [redacted]
45. Deposition and Exhibits of [redacted]
46. Deposition and Exhibits of [redacted]
47. Deposition and Exhibits of [redacted]
48. Deposition and Exhibits of [redacted]
49. Deposition and Exhibits of [redacted]
50. Deposition and Exhibits of [redacted]
51. Deposition and Exhibits of [redacted]

52. Deposition and Exhibits of South Carolina (Rebecca Hartner), April 23, 2024
53. Deposition and Exhibits of Indiana (Jamie Weber), April 26, 2024
54. Deposition and Exhibits of Indiana (Steven Taterka), April 26, 2024
55. Deposition and Exhibits of Nevada (Lucas Tucker), April 29, 2024
56. Deposition and Exhibits of Arkansas (Chuck Harder), May 1, 2024
57. Deposition and Exhibits of Alaska (Jeff Pickett), May 3, 2024
58. Deposition and Exhibits of Florida (Andrew Butler), April 22, 2024
59. Deposition and Exhibits of Idaho (John Olson), May 3, 2024
60. Deposition and Exhibits of Idaho (Stephanie Guyon), May 3, 2024
61. Deposition and Exhibits of Kentucky (Jonathan Farmer), April 25, 2024
62. Deposition and Exhibits of Louisiana (Patrick Voelker), May 3, 2024
63. Deposition and Exhibits of Mississippi (Crystal Secoy), April 25, 2024

64. Deposition and Exhibits of Mississippi (Sid Salter), April 25, 2024
65. Deposition and Exhibits of Missouri (Michael Schwalbert), May 10, 2024
66. Deposition and Exhibits of Montana (Anna Schneider), May 1, 2024
67. Deposition and Exhibits of North Dakota (Elin Alm), May 2, 2024
68. Deposition and Exhibits of Puerto Rico (Guarinonex Diaz Martinez), May 1, 2024
69. Deposition and Exhibits of South Dakota (Jonathan Van Patten), April 29, 2024
70. Deposition and Exhibits of Texas (Trevor Young), May 24, 2024
71. Deposition and Exhibits of Texas (Justin Gordon), April 17, 2024
72. Deposition and Exhibits of Utah (Marie Martin), April 30, 2024
73. Deposition and Exhibits of Utah (Melanie Hall), April 30, 2024
74. Deposition and Exhibits of [redacted]
75. Deposition and Exhibits of [redacted]
76. Deposition and Exhibits of [redacted]
77. Deposition and Exhibits of [redacted]
78. Deposition and Exhibits of [redacted]
79. Deposition and Exhibits of [redacted]
80. Deposition and Exhibits of [redacted]
81. Deposition and Exhibits of [redacted]
82. Deposition and Exhibits of [redacted]
83. Deposition and Exhibits of [redacted]
84. Deposition and Exhibits of [redacted]

All available deposition transcripts and exhibits within the matter of *USA v. Google*, Case Number: 1:23-cv-00108-LMB-JFA, including:

85. Deposition and Exhibits of [redacted]
86. Deposition and Exhibits of [redacted]
87. Deposition and Exhibits of [redacted]
88. Deposition and Exhibits of [redacted]
89. Deposition and Exhibits of [redacted]
90. Deposition and Exhibits of [redacted]
91. Deposition and Exhibits of [redacted]
92. Deposition and Exhibits of [redacted]
93. Deposition and Exhibits of [redacted]
94. Deposition and Exhibits of [redacted]
95. Deposition and Exhibits of [redacted]
96. Deposition and Exhibits of [redacted]

97. Deposition and Exhibits of [redacted]
98. Deposition and Exhibits of [redacted]
99. Deposition and Exhibits of [redacted]
100. Deposition and Exhibits of [redacted]
101. Deposition and Exhibits of [redacted]
102. Deposition and Exhibits of [redacted]
103. Deposition and Exhibits of [redacted]
104. Deposition and Exhibits of [redacted]
105. Deposition and Exhibits of [redacted]
106. Deposition and Exhibits of [redacted]
107. Deposition and Exhibits of [redacted]
108. Deposition and Exhibits of [redacted]
109. Deposition and Exhibits of [redacted]

110. Deposition and Exhibits of [redacted] (November, 11, 2023)
111. Deposition and Exhibits of [redacted] (August 15, 2023)
112. Deposition and Exhibits of [redacted] (November 14, 2023)
113. Deposition and Exhibits of [redacted] (November 15, 2023)
114. Deposition and Exhibits of [redacted] (November 14, 2023)
115. Deposition and Exhibits of [redacted] (30B6 errata only) (November 14, 2023)
116. Deposition and Exhibits of [redacted] (November 3, 2023)
117. Deposition and Exhibits of [redacted] (August 16, 2023)
118. Deposition and Exhibits of [redacted] (November 7, 2023)
119. Deposition and Exhibits of [redacted] (November 9, 2023)
120. Deposition and Exhibits of [redacted] (October 30, 2023)
121. Deposition and Exhibits of [redacted] (August 11, 2023)
122. Deposition and Exhibits of [redacted] (November 2, 2023)
123. Deposition and Exhibits of [redacted] (November 16, 2023)
124. Deposition and Exhibits of [redacted] (August 29, 2023)
125. Deposition and Exhibits of [redacted] (November 14-15, 2023)
126. Deposition and Exhibits of [redacted] (April 1, 2024)
127. Deposition and Exhibits of [redacted] (November 3, 2024)
128. Deposition and Exhibits of [redacted] (November 3, 2024)
129. Deposition and Exhibits of [redacted] (30(b)6) (November 14, 2023)
130. Deposition and Exhibits of [redacted] (August 16, 2023)
131. Deposition and Exhibits of [redacted] (November 7, 2023)
132. Deposition and Exhibits of [redacted] (November 9, 2023)
133. Deposition and Exhibits of [redacted] (April 3, 2024)
134. Deposition and Exhibits of [redacted] (October 10, 2023 and November 8, 2023)
135. Deposition and Exhibits of [redacted] (April 17, 2024)
136. Deposition and Exhibits of [redacted] (April 29, 2024)
137. Deposition and Exhibits of [redacted] (Noveber 11, 2023)
138. Deposition and Exhibits of [redacted] (October 10, 2023)

All available deposition transcripts and exhibits within the matter of *In re: Google Digital Advertising Antitrust Litigation*, Case Number: 1:21-md-03010-PKC, including the depositions and exhibits of:

139.	[redacted]	6/19/2024
140.	[redacted]	6/20/2024
141.	[redacted]	6/21/2024
142.	[redacted]	5/21/2024
143.	[redacted]	6/25/2024
144.	[redacted]	6/25/2024
145.	[redacted]	6/27/2024
146.	[redacted]	7/23/2024
147.	[redacted]	7/23/2024
148.	[redacted]	6/18/2024
149.	[redacted]	5/7/2024
150.	[redacted]	7/9/2024
151.	[redacted]	7/10/2024
152.	[redacted]	4/25/2024
153.	[redacted]	7/10/2024
154.	[redacted]	6/24/2024
155.	[redacted]	7/12/2024
156.	[redacted]	6/12/2024
157.	[redacted]	6/13/2024
158.	[redacted]	5/2/2024
159.	[redacted]	6/28/2024
160.	[redacted]	6/6/2024
161.	[redacted]	6/28/2024
162.	[redacted]	7/3/2024
163.	[redacted]	6/4/2024
164.	[redacted]	7/28/2024
165.	[redacted]	7/10/2024
166.	[redacted]	6/25/2024
167.	[redacted]	6/26/2024
168.	[redacted]	6/10/2024
169.	[redacted]	6/27/2024
170.	[redacted]	6/13/2024
171.	[redacted]	6/7/2024
172.	[redacted]	6/25/2024
173.	[redacted]	6/28/2024
174.	[redacted]	5/24/2024
175.	[redacted]	6/24/2024
176.	[redacted]	6/27/2024
177.	[redacted]	6/11/2024

178.	[redacted]	6/12/2024

Other available deposition transcripts and exhibits, including the depositions and exhibits of:

179.	[redacted]	10/2/2020
180.	[redacted]	10/16/2020
181.	[redacted]	7/28/2020
182.	[redacted]	7/21/2020
183.	[redacted]	10/26/2020
184.	[redacted]	11/6/2020
185.	[redacted]	7/31/2020
186.	[redacted]	9/25/2020
187.	[redacted]	10/20/2020
188.	[redacted]	7/17/2020
189.	[redacted]	11/9/2020
190.	[redacted]	11/19/2020
191.	[redacted]	7/24/2020
192.	[redacted]	7/14/2020
193.	[redacted]	11/10/2020
194.	[redacted]	11/2/2020
195.	[redacted]	9/28/2020
196.	[redacted]	2/3/2022
197.	[redacted]	8/11/2021
198.	[redacted]	2/28/2022
199.	[redacted]	10/19/2021
200.	[redacted]	12/9/2021
201.	[redacted]	9/17/2021
202.	[redacted]	11/20/2020
203.	[redacted]	3/30/2021
204.	[redacted]	10/28/2021
205.	[redacted]	8/10/2021
206.	[redacted]	3/31/2021
207.	[redacted]	4/2/2021
208.	[redacted]	4/22/2021
209.	[redacted]	10/28/2021
210.	[redacted]	7/22/2021
211.	[redacted]	10/6/2021
212.	[redacted]	7/20/2021
213.	[redacted]	8/12/2021
214.	[redacted]	9/28/2021
215.	[redacted]	5/17/2021
216.	[redacted]	9/7/2021

D. Expert Reports & Declarations

All available expert reports, including appendices, backup materials, and cited materials, within the matter of *The State of Texas, et al. v. Google*, Case Number: 4:20-cv-00957-SDJ, including:

1. 2024.06.07 Expert Report of Jeffrey S. Andrien
2. 2024.06.07 Expert Report of Joshua Gans, as well as 2024.07.24 Errata and Supplemental Appendix D
3. 2024.06.07 Expert Report of Jacob Hostetler
4. 2024.06.07 Expert Report of John Chandler
5. 2024.06.07 Expert Report of Matthew Weinberg
6. 2024.06.07 Expert Report of Parag Pathak
7. 2024.07.30 Expert Report of Anindya Ghose
8. 2024.07.30 Expert Report of Donna L. Hoffman
9. 2024.07.30 Expert Report of Douglas Skinner
10. 2024.07.30 Expert Report of Itamar Simonson
11. 2024.07.30 Expert Report of Martin C. Rinard
12. 2024.07.30 Expert Report of Paul R. Milgrom
13. 2024.07.30 Expert Report of Steven N. Wiggins
14. 2024.08.06 Expert Report of Michael R. Baye
15. 2024.08.06 Expert Report of Jason Nieh

All available expert reports (with redactions) within the matter of *USA v. Google,* Case Number: 1:23-cv-00108-LMB-JFA, including:

1. Declarations of Google Employees
2. 2023.12.22 Expert Report of Gabriel Weintraub, GOOG-AT-MDL-C-000018734
3. 2023.12.22 Expert Report of R. Ravi, GOOG-AT-MDL-C-000019017
4. 2023.12.22 Expert Report of Robin S. Lee, GOOG-AT-MDL-C-000019273
5. 2023.12.22 Expert Report of Rosa Abrantes-Metz, GOOG-AT-MDL-C-000019786
6. 2023.12.22 Expert Report of Thomas S. Respess, GOOG-AT-MDL-C-000020106
7. 2023.12.22 Expert Report of Timothy Simcoe, GOOG-AT-MDL-C-000020274
8. 2024.01.13 Errata to Abrantes-Metz Expert Report, GOOG-AT-MDL-C-000020435
9. 2024.01.13 Errata to Ravi Expert Report, GOOG-AT-MDL-C-000020437
10. 2024.01.13 Errata to Respess Expert Report, GOOG-AT-MDL-C-000020440
11. 2024.01.13 Errata to Simcoe Expert Report, GOOG-AT-MDL-C-000020467
12. 2024.01.13 Errata to Weintraub Expert Report, GOOG-AT-MDL-C-000020471
13. 2024.01.23 Chevalier Expert Report, GOOG-AT-MDL-C-000020474

14. 2024.01.23 Ferrante Expert Report, GOOG-AT-MDL-C-000020714
15. 2024.01.23 Ghose Expert Report, GOOG-AT-MDL-C-000020767
16. 2024.01.23 Israel Expert Report, GOOG-AT-MDL-C-000021036
17. 2024.01.23 Milgrom Expert Report, GOOG-AT-MDL-C-000021794
18. 2024.01.23 Rinard Expert Report, GOOG-AT-MDL-C-000022191
19. 2024.01.23 Shirky Expert Report, GOOG-AT-MDL-C-000022229
20. 2024.01.23 Simonson Expert Report, GOOG-AT-MDL-C-000022290
21. 2024.01.23 Skinner Expert Report, GOOG-AT-MDL-C-000022948
22. 2024.02.13 Expert Rebuttal Report of Adoria Lim, GOOG-AT-MDL-C-000023002
23. 2024.02.13 Expert Rebuttal Report of Gabriel Weintraub, GOOG-AT-MDL-C-000023226
24. 2024.02.13 Expert Rebuttal Report of Kenneth Wilbur, GOOG-AT-MDL-C-000023322
25. 2024.02.13 Expert Rebuttal Report of R. Ravi, GOOG-AT-MDL-C-000023435
26. 2024.02.13 Expert Rebuttal Report of Robin S. Lee, GOOG-AT-MDL-C-000023516
27. 2024.02.13 Expert Rebuttal Report of Rosa Abrantes-Metz, GOOG-AT-MDL-C-000023887
28. 2024.02.13 Expert Rebuttal Report of Timothy Simcoe, GOOG-AT-MDL-C-000024064
29. 2024.02.13 Expert Rebuttal Report of Wayne Hoyer, GOOG-AT-MDL-C-000024138
30. 2024.02.13 Expert Rebuttal Report of Wenke Lee, GOOG-AT-MDL-C-000024270
31. 2024.02.16 Errata to Ravi Rebuttal Report, GOOG-AT-MDL-C-000024387
32. 2024.02.20 Errata to Simcoe Rebuttal Report, GOOG-AT-MDL-C-000024389
33. 2024.02.23 Errata to Weintraub Rebuttal Report, GOOG-AT-MDL-C-000024390
34. 2024.02.23 Supplemental Errata to Weintraub Expert Report, GOOG-AT-MDL-C-000024391
35. 2024.02.24 Errata to Wilbur Rebuttal Report, GOOG-AT-MDL-C-000024392
36. 2024.02.26 Errata to Hoyer Rebuttal Report, GOOG-AT-MDL-C-000024397
37. 2024.02.28 Errata to Abrantes-Metz Rebuttal Report, GOOG-AT-MDL-C-000024399
38. 2024.03.04 Expert Supplemental Report of Robin S. Lee, GOOG-AT-MDL-C-000024403
39. 2024.03.08 Consolidated Errata to Lee Rebuttal Report, GOOG-AT-MDL-C-000024436
40. 2024.01.13 Expert Report of Weintraub Errata, GOOG-AT-MDL-C-000040965
41. 2024.01.13 Expert Report of Simcoe Errata, GOOG-AT-MDL-C-000040961
42. 2024.01.13 Expert Report of Respess Errata_with Figure Errata_Redacted, GOOG-AT-MDL-C-000040934
43. 2024.01.13 Expert Report of R Ravi Errata, GOOG-AT-MDL-C-000040931
44. 2024.01.13 Expert Report of Abrantes-Metz Errata, GOOG-AT-MDL-C-000040929
45. 2024.03.08 Consolidated Errata to Lee Rebuttal Report, GOOG-AT-MDL-C-000040926

46. 2024.03.04 Expert Supplemental Report of Robin S. Lee, PhD, GOOG-AT-MDL-C-000040893
47. 2024.02.28 Rebuttal Report Errata of Rosa Abrantes-Metz Signed, GOOG-AT-MDL-C-000040889
48. 2024.02.25 Expert Rebuttal Report of Hoyer Errata, GOOG-AT-MDL-C-000040887
49. 2024.02.24 Wilbur Rebuttal Errata, GOOG-AT-MDL-C-000040882
50. 2024.02.23 Weintraub Rebuttal Report Errata, GOOG-AT-MDL-C-000040881
51. 2024.02.23 Expert Report of Weintraub Supplemental Errata, GOOG-AT-MDL-C-000040880
52. 2024.02.20 Errata to Simcoe Rebuttal Report, GOOG-AT-MDL-C-000040879
53. 2024.02.16 Errata to Ravi Rebuttal Report (Highly Confidential), GOOG-AT-MDL-C-000040877
54. 2024.02.13 Rebuttal Report of Rosa Abrantes-Metz, GOOG-AT-MDL-C-000040700
55. 2024.02.13 Expert Report of Wenke Lee, GOOG-AT-MDL-C-000040583
56. 2024.02.13 Expert Rebuttal Report of Wayne Hoyer, GOOG-AT-MDL-C-000040451
57. 2024.02.13 Expert Rebuttal Report of Timothy Simcoe_Redacted, GOOG-AT-MDL-C-000040377
58. 2024.02.13 Expert Rebuttal Report of Robin S. Lee_Redacted, GOOG-AT-MDL-C-000040006
59. 2024.02.13 Expert Rebuttal Report of R Ravi, GOOG-AT-MDL-C-000039925
60. 2024.02.13 Expert Rebuttal Report of Kenneth Wilbur_Redacted, GOOG-AT-MDL-C-000039812
61. 2024.02.13 Expert Rebuttal Report of Gabriel Weintraub_Redacted, GOOG-AT-MDL-C-000039716
62. 2024.02.13 Expert Rebuttal Report of Adoria Lim_Redacted, GOOG-AT-MDL-C-000039492
63. 2024.01.23 Expert Report of William Clay Shirky, GOOG-AT-MDL-C-000039431
64. 2024.01.23 Expert Report of Paul R. Milgrom, GOOG-AT-MDL-C-000039034
65. 2024.01.23 Expert Report of Martin C. Rinard, GOOG-AT-MDL-C-000038996
66. 2024.01.23 Expert Report of Mark A. Israel_Redacted, GOOG-AT-MDL-C-000038238
67. 2024.01.23 Expert Report of Judith A. Chevalier_Redacted, GOOG-AT-MDL-C-000037998
68. 2024.01.23 Expert Report of Itamar Simonson, GOOG-AT-MDL-C-000037340
69. 2024.01.23 Expert Report of Douglas Skinner, GOOG-AT-MDL-C-000037286
70. 2024.01.23 Expert Report of Anthony J. Ferrante, GOOG-AT-MDL-C-000037233
71. 2024.01.23 Expert Report of Anindya Ghose_Redacted, GOOG-AT-MDL-C-000036954
72. 2023.12.22 Expert Report of Timothy Simcoe_Redacted, GOOG-AT-MDL-C-000036793
73. 2023.12.22 Expert Report of Thomas Respess_Redacted, GOOG-AT-MDL-C-000036625
74. 2023.12.22 Expert Report of Rosa Abrantes-Metz_Redacted, GOOG-AT-MDL-C-000036305

75. 2023.12.22 Expert Report of Robin S. Lee, PhD_Redacted, GOOG-AT-MDL-C-000035792
76. 2023.12.22 Expert Report of R Ravi_Redacted, GOOG-AT-MDL-C-000035536
77. 2023.12.22 Expert Report of Gabriel Weintraub_Redacted, GOOG-AT-MDL-C-000035253

E. Bates Stamped Productions, including access to Plaintiffs' entire production database, as well as the following documents and Google and third-party productions made since June 7, 2024

1. [redacted]
2. [redacted]
3. [redacted]
4. [redacted]
5. [redacted]
6. [redacted]
7. [redacted]
8. [redacted]
9. [redacted]
10. [redacted]
11. [redacted]
12. [redacted]
13. [redacted]
14. [redacted]
15. [redacted]
16. [redacted]
17. [redacted]
18. [redacted]
19. [redacted]
20. [redacted]
21. [redacted]
22. [redacted]
23. [redacted]
24. [redacted]
25. [redacted]
26. [redacted]
27. [redacted]
28. [redacted]
29. [redacted]
30. [redacted]
31. [redacted]
32. [redacted]
33. [redacted]
34. [redacted]
35. [redacted]
36. [redacted]
37. [redacted]
38. [redacted]
39. [redacted]
40. GOOG-AT-MDL-001263607
41. GOOG-AT-MDL-001390730
42. GOOG-AT-MDL-001391213
43. GOOG-AT-MDL-001933227
44. GOOG-AT-MDL-002105969
45. GOOG-AT-MDL-002105984
46. GOOG-AT-MDL-002124829
47. GOOG-AT-MDL-002390899
48. GOOG-AT-MDL-002393442
49. GOOG-AT-MDL-003161451
50. GOOG-AT-MDL-004074544
51. GOOG-AT-MDL-004232880
52. GOOG-AT-MDL-004233138
53. GOOG-AT-MDL-004300268
54. GOOG-AT-MDL-004416785
55. GOOG-AT-MDL-004436768
56. GOOG-AT-MDL-004555181
57. GOOG-AT-MDL-006099844
58. GOOG-AT-MDL-006161050
59. GOOG-AT-MDL-006334729
60. GOOG-AT-MDL-006873424
61. GOOG-AT-MDL-006966530
62. GOOG-AT-MDL-007175167
63. GOOG-AT-MDL-007343585
64. GOOG-AT-MDL-007346556
65. GOOG-AT-MDL-007364833
66. GOOG-AT-MDL-007375672
67. GOOG-AT-MDL-007387750
68. GOOG-AT-MDL-007397182
69. GOOG-AT-MDL-007397197
70. GOOG-AT-MDL-008148533 / GOOG-AT-MDL-008148529
71. GOOG-AT-MDL-008517788
72. GOOG-AT-MDL-008588684
73. GOOG-AT-MDL-008682082 / GOOG-AT-MDL-008682071
74. GOOG-AT-MDL-008754374

75. GOOG-AT-MDL-008835346
76. GOOG-AT-MDL-008858602
77. GOOG-AT-MDL-008881206
78. GOOG-AT-MDL-008886980
79. GOOG-AT-MDL-008953893
80. GOOG-AT-MDL-008964888
81. GOOG-AT-MDL-008979664
82. GOOG-AT-MDL-008991390
83. GOOG-AT-MDL-009026140
84. GOOG-AT-MDL-009289718
85. GOOG-AT-MDL-009291120
86. GOOG-AT-MDL-009299907
87. GOOG-AT-MDL-009321580
88. GOOG-AT-MDL-009429957
89. GOOG-AT-MDL-012512067
90. GOOG-AT-MDL-012514705
91. GOOG-AT-MDL-012524006
92. GOOG-AT-MDL-012549335
93. GOOG-AT-MDL-012693796
94. GOOG-AT-MDL-012767138
95. GOOG-AT-MDL-012837016
96. GOOG-AT-MDL-012857198
97. GOOG-AT-MDL-013290688
98. GOOG-AT-MDL-013291089
99. GOOG-AT-MDL-013292974
100. GOOG-AT-MDL-013299524
101. GOOG-AT-MDL-013299531
102. GOOG-AT-MDL-013300202
103. GOOG-AT-MDL-013378392
104. GOOG-AT-MDL-013908958
105. GOOG-AT-MDL-013918668
106. GOOG-AT-MDL-014427012
107. GOOG-AT-MDL-014460206
108. GOOG-AT-MDL-014462378
109. GOOG-AT-MDL-014486274
110. GOOG-AT-MDL-014524447
111. GOOG-AT-MDL-014618288
112. GOOG-AT-MDL-015241235
113. GOOG-AT-MDL-015622194
114. GOOG-AT-MDL-015844174
115. GOOG-AT-MDL-015929587
116. GOOG-AT-MDL-015997353
117. GOOG-AT-MDL-016457027
118. GOOG-AT-MDL-016534880
119. GOOG-AT-MDL-016627159
120. GOOG-AT-MDL-016656237
121. GOOG-AT-MDL-016772599
122. GOOG-AT-MDL-016838311
123. GOOG-AT-MDL-016924839
124. GOOG-AT-MDL-016937590
125. GOOG-AT-MDL-016943922
126. GOOG-AT-MDL-016967094
127. GOOG-AT-MDL-017187837
128. GOOG-AT-MDL-017394050
129. GOOG-AT-MDL-017494582
130. GOOG-AT-MDL-017664768
131. GOOG-AT-MDL-017746412
132. GOOG-AT-MDL-017749638
133. GOOG-AT-MDL-017762649
134. GOOG-AT-MDL-017864022
135. GOOG-AT-MDL-018248228
136. GOOG-AT-MDL-018427318
137. GOOG-AT-MDL-018448707
138. GOOG-AT-MDL-018548592
139. GOOG-AT-MDL-018618351
140. GOOG-AT-MDL-018652651
141. GOOG-AT-MDL-018998910
142. GOOG-AT-MDL-019001498
143. GOOG-AT-MDL-019306356
144. GOOG-AT-MDL-019386250
145. GOOG-AT-MDL-019552139
146. GOOG-AT-MDL-019571201
147. GOOG-AT-MDL-019588187
148. GOOG-AT-MDL-019633443
149. GOOG-AT-MDL-019642313
150. GOOG-AT-MDL-019653406
151. GOOG-AT-MDL-019721340
152. GOOG-AT-MDL-019767203
153. GOOG-AT-MDL-B-000134141
154. GOOG-AT-MDL-B-001084151
155. GOOG-AT-MDL-B-001140202
156. GOOG-AT-MDL-B-002087955
157. GOOG-AT-MDL-B-002088697
158. GOOG-AT-MDL-B-002088752
159. GOOG-AT-MDL-B-002088926
160. GOOG-AT-MDL-B-002090567
161. GOOG-AT-MDL-B-002091565
162. GOOG-AT-MDL-B-002095353
163. GOOG-AT-MDL-B-002095501
164. GOOG-AT-MDL-B-002095769
165. GOOG-AT-MDL-B-002097533
166. GOOG-AT-MDL-B-002097570

167. GOOG-AT-MDL-B-002097648
168. GOOG-AT-MDL-B-002098265
169. GOOG-AT-MDL-B-002099366
170. GOOG-AT-MDL-B-002105135
171. GOOG-AT-MDL-B-002500395
172. GOOG-AT-MDL-B-002514153
173. GOOG-AT-MDL-B-002547489
174. GOOG-AT-MDL-B-002552122
175. GOOG-AT-MDL-B-002624643
176. GOOG-AT-MDL-B-002760309
177. GOOG-AT-MDL-B-002762758
178. GOOG-AT-MDL-B-002763194
179. GOOG-AT-MDL-B-002764178
180. GOOG-AT-MDL-B-002764191
181. GOOG-AT-MDL-B-002797051
182. GOOG-AT-MDL-B-002803919
183. GOOG-AT-MDL-B-002837096
184. GOOG-AT-MDL-B-003181628
185. GOOG-AT-MDL-B-003735823
186. GOOG-AT-MDL-B-003741154
187. GOOG-AT-MDL-B-004008753
188. GOOG-AT-MDL-B-004015880
189. GOOG-AT-MDL-B-004016318
190. GOOG-AT-MDL-B-004243337
191. GOOG-AT-MDL-B-004247242
192. GOOG-AT-MDL-B-004265772
193. GOOG-AT-MDL-B-004425247
194. GOOG-AT-MDL-B-004680051
195. GOOG-AT-MDL-B-005083974
196. GOOG-AT-MDL-B-005167304
197. GOOG-AT-MDL-B-005168118
198. GOOG-AT-MDL-B-005170475
199. GOOG-AT-MDL-B-005282318
200. GOOG-AT-MDL-B-005372599
201. GOOG-AT-MDL-B-005457387
202. GOOG-AT-MDL-B-006069467
203. GOOG-AT-MDL-B-006316352
204. GOOG-AT-MDL-B-006365895
205. GOOG-AT-MDL-B-006365981
206. GOOG-AT-MDL-B-006939056
207. GOOG-AT-MDL-B-007212533
208. GOOG-AT-MDL-B-007229334
209. GOOG-AT-MDL-B-007232867
210. GOOG-AT-MDL-B-007353902
211. GOOG-DOJ-12948968
212. GOOG-DOJ-13897780
213. GOOG-DOJ-13899823
214. GOOG-DOJ-13911836
215. GOOG-DOJ-13930748
216. GOOG-DOJ-13940086
217. GOOG-DOJ-14008698
218. GOOG-DOJ-14034714
219. GOOG-DOJ-14113270
220. GOOG-DOJ-14139857
221. GOOG-DOJ-14155066
222. GOOG-DOJ-14156827
223. GOOG-DOJ-14161619
224. GOOG-DOJ-14161943
225. GOOG-DOJ-14232497
226. GOOG-DOJ-14352302
227. GOOG-DOJ-14352774
228. GOOG-DOJ-14365517
229. GOOG-DOJ-14433486
230. GOOG-DOJ-14433633
231. GOOG-DOJ-14435110
232. GOOG-DOJ-14436029
233. GOOG-DOJ-14453674
234. GOOG-DOJ-14458088
235. GOOG-DOJ-14494204
236. GOOG-DOJ-14544715
237. GOOG-DOJ-14544743
238. GOOG-DOJ-14556699
239. GOOG-DOJ-14712739
240. GOOG-DOJ-14717683
241. GOOG-DOJ-14733660
242. GOOG-DOJ-14735212
243. GOOG-DOJ-14735427
244. GOOG-DOJ-14740782
245. GOOG-DOJ-14743636
246. GOOG-DOJ-14870370
247. GOOG-DOJ-15021111
248. GOOG-DOJ-15022600
249. GOOG-DOJ-15022611
250. GOOG-DOJ-15029681
251. GOOG-DOJ-15041717
252. GOOG-DOJ-15042589
253. GOOG-DOJ-15064786
254. GOOG-DOJ-15071642
255. GOOG-DOJ-15073261
256. GOOG-DOJ-15076754
257. GOOG-DOJ-15140608
258. GOOG-DOJ-15173140

259. GOOG-DOJ-15186471
260. GOOG-DOJ-15278949
261. GOOG-DOJ-15371972
262. GOOG-DOJ-15417170
263. GOOG-DOJ-15426837
264. GOOG-DOJ-15427800
265. GOOG-DOJ-15428154
266. GOOG-DOJ-15432090
267. GOOG-DOJ-15433127
268. GOOG-DOJ-15434544
269. GOOG-DOJ-15445478
270. GOOG-DOJ-15526215
271. GOOG-DOJ-15590044
272. GOOG-DOJ-15595543
273. GOOG-DOJ-15597407
274. GOOG-DOJ-15598151
275. GOOG-DOJ-15601229
276. GOOG-DOJ-15616526
277. GOOG-DOJ-15621021
278. GOOG-DOJ-15628088
279. GOOG-DOJ-15631978
280. GOOG-DOJ-15637110
281. GOOG-DOJ-15675958
282. GOOG-DOJ-15733030
283. GOOG-DOJ-15766965
284. GOOG-DOJ-15789579
285. GOOG-DOJ-16271009
286. GOOG-DOJ-16732920
287. GOOG-DOJ-16976651
288. GOOG-DOJ-17101302
289. GOOG-DOJ-17105055
290. GOOG-DOJ-17581278
291. GOOG-DOJ-19090482
292. GOOG-DOJ-19119112
293. GOOG-DOJ-27760500_DUP_1
294. GOOG-DOJ-27762649
295. GOOG-DOJ-27762691
296. GOOG-DOJ-27796586
297. GOOG-DOJ-27803156
298. GOOG-DOJ-27803505
299. GOOG-DOJ-27803533
300. GOOG-DOJ-27804205
301. GOOG-DOJ-27804876
302. GOOG-DOJ-28385887
303. GOOG-DOJ-28420330
304. GOOG-DOJ-29427368
305. GOOG-DOJ-29453844
306. GOOG-DOJ-30066802
307. GOOG-DOJ-32018179
308. GOOG-DOJ-32018452
309. GOOG-DOJ-32022422
310. GOOG-DOJ-32062262
311. GOOG-DOJ-32262980
312. GOOG-DOJ-32265694
313. GOOG-DOJ-AT-00003565
314. GOOG-DOJ-AT-00087809
315. GOOG-DOJ-AT-00252559
316. GOOG-DOJ-AT-00569953
317. GOOG-DOJ-AT-00575435
318. GOOG-DOJ-AT-00588995
319. GOOG-DOJ-AT-00651561
320. GOOG-DOJ-AT-00663174
321. GOOG-DOJ-AT-00750354
322. GOOG-DOJ-AT-01019411
323. GOOG-DOJ-AT-01027937
324. GOOG-DOJ-AT-01514374
325. GOOG-DOJ-AT-01525829
326. GOOG-DOJ-AT-01682499
327. GOOG-DOJ-AT-01811246
328. GOOG-DOJ-AT-01814428
329. GOOG-DOJ-AT-01814688
330. GOOG-DOJ-AT-01815482
331. GOOG-DOJ-AT-01816686
332. GOOG-DOJ-AT-01848255
333. GOOG-DOJ-AT-01933085
334. GOOG-DOJ-AT-02070545
335. GOOG-DOJ-AT-02122748
336. GOOG-DOJ-AT-02146896
337. GOOG-DOJ-AT-02153612
338. GOOG-DOJ-AT-02275074
339. GOOG-DOJ-AT-02275933
340. GOOG-DOJ-AT-02319393
341. GOOG-DOJ-AT-02482767
342. GOOG-DOJ-AT-02634336
343. GOOG-DOJ-AT-02639830
344. GOOG-NE-01787563
345. GOOG-NE-02557667
346. GOOG-NE-02558055
347. GOOG-NE-03231727
348. GOOG-NE-03236272
349. GOOG-NE-03467508
350. GOOG-NE-03616222

351. GOOG-NE-03627597
352. GOOG-NE-03727939
353. GOOG-NE-03844060
354. GOOG-NE-03869994
355. GOOG-NE-03872605
356. GOOG-NE-04304005
357. GOOG-NE-04689402
358. GOOG-NE-05241137
359. GOOG-NE-05270679
360. GOOG-NE-06113062
361. GOOG-NE-06547825
362. GOOG-NE-06568562
363. GOOG-NE-06568719
364. GOOG-NE-06729717
365. GOOG-NE-06841582
366. GOOG-NE-07841238
367. GOOG-NE-07946731
368. GOOG-NE-08116078
369. GOOG-NE-08112779
370. GOOG-NE-09138394
371. GOOG-NE-09698556
372. GOOG-NE-09711564
373. GOOG-NE-10573952
374. GOOG-NE-10660658
375. GOOG-NE-10660882
376. GOOG-NE-10777158
377. GOOG-NE-10952313
378. GOOG-NE-10976657
379. GOOG-NE-10977717
380. GOOG-NE-10978402
381. GOOG-NE-10985565
382. GOOG-NE-11813912
383. GOOG-NE-11847370
384. GOOG-NE-11877282
385. GOOG-NE-12727978
386. GOOG-NE-12738745
387. GOOG-NE-12742008
388. GOOG-NE-12787759
389. GOOG-NE-12885078
390. GOOG-NE-13004102
391. GOOG-NE-13200100
392. GOOG-NE-13205235
393. GOOG-NE-13205865
394. GOOG-NE-13210272
395. GOOG-NE-13327192
396. GOOG-NE-13340752
397. GOOG-NE-13349343
398. GOOG-NE-13367768
399. GOOG-NE-13369624
400. GOOG-NE-13373170
401. GOOG-NE-13374150
402. GOOG-NE-13379438
403. GOOG-NE-13386334
404. GOOG-NE-13389481
405. GOOG-NE-13390045
406. GOOG-NE-13393115
407. GOOG-NE-13401911
408. GOOG-NE-13405025
409. GOOG-NE-13415537
410. GOOG-NE-13541543
411. GOOG-NE-13569193
412. GOOG-NE-13577462
413. GOOG-NE-13589899
414. GOOG-NE-13624783
415. GOOG-TEX-00039266
416. GOOG-TEX-00079344
417. GOOG-TEX-00089835
418. GOOG-TEX-00090151
419. GOOG-TEX-00090282
420. GOOG-TEX-00090969
421. GOOG-TEX-00092657
422. GOOG-TEX-00105202
423. GOOG-TEX-00105361
424. GOOG-TEX-00106255
425. GOOG-TEX-00109167
426. GOOG-TEX-00110068
427. GOOG-TEX-00110540
428. GOOG-TEX-00110544
429. GOOG-TEX-00111494
430. GOOG-TEX-00116069
431. GOOG-TEX-00116639
432. GOOG-TEX-00119737
433. GOOG-TEX-00119845
434. GOOG-TEX-00120626
435. GOOG-TEX-00120775
436. GOOG-TEX-00156142
437. GOOG-TEX-00167119
438. GOOG-TEX-00177559
439. GOOG-TEX-00206687
440. GOOG-TEX-00216163
441. GOOG-TEX-00234150
442. GOOG-TEX-00240572

443. GOOG-TEX-00270127
444. GOOG-TEX-00271177
445. GOOG-TEX-00309326
446. GOOG-TEX-00326649
447. GOOG-TEX-00336527
448. GOOG-TEX-00344083
449. GOOG-TEX-00370306
450. GOOG-TEX-00374779
451. GOOG-TEX-00375239
452. GOOG-TEX-00452866
453. GOOG-TEX-00513684
454. GOOG-TEX-00643890
455. GOOG-TEX-00689539
456. GOOG-TEX-00705131
457. GOOG-TEX-00715805
458. GOOG-TEX-00716782
459. GOOG-TEX-00777573
460. GOOG-TEX-00778301
461. GOOG-TEX-00797340
462. GOOG-TEX-00806640
463. GOOG-TEX-00806682
464. GOOG-TEX-00814407
465. GOOG-TEX-00825713
466. GOOG-TEX-00828547
467. GOOG-TEX-00831660
468. GOOG-TEX-00850729
469. GOOG-TEX-00858576
470. GOOG-TEX-00905673
471. GOOG-TEX-00959457
472. GOOG-TEX-00959461
473. GOOG-TEX-00969525
474. GOOG-TEX-00978814
475. GOOG-TEX-01004466
476. GOOG-TEX-01036150
477. GOOG-TEX-01142635
478. GOOG-TEX-01244428
479. GOOG-TEX-01279945
480. GOOG-TEX-00974499
481. GOOG-TEX-00001418
482. GOOG-TEX-00597317
483. [redacted]
484.
485.
486.
487.
488.
489. [redacted]
490.
491.
492.
493.
494.
495.
496.
497.
498.
499.
500.
501.
502.
503.
504.
505.
506.
507.
508.
509.
510.
511. GOOG-AT-MDL-000991808
512. GOOG-AT-MDL-001386659
513. GOOG-AT-MDL-001396947
514. GOOG-AT-MDL-001398467
515. GOOG-AT-MDL-001403170
516. GOOG-AT-MDL-001412455
517. GOOG-AT-MDL-001428366
518. GOOG-AT-MDL-001439138
519. GOOG-AT-MDL-001446832
520. GOOG-AT-MDL-001828386
521. GOOG-AT-MDL-001972093
522. GOOG-AT-MDL-002468282
523. GOOG-AT-MDL-002468549
524. GOOG-AT-MDL-003408566
525. GOOG-AT-MDL-003518690
526. GOOG-AT-MDL-003530493
527. GOOG-AT-MDL-003596410
528. GOOG-AT-MDL-003664784
529. GOOG-AT-MDL-003675001
530. GOOG-AT-MDL-003752136
531. GOOG-AT-MDL-003832428
532. GOOG-AT-MDL-003960772
533. GOOG-AT-MDL-003960884
534. GOOG-AT-MDL-003984934

535. GOOG-AT-MDL-004284828
536. GOOG-AT-MDL-004285638
537. GOOG-AT-MDL-004324549
538. GOOG-AT-MDL-004416737
539. GOOG-AT-MDL-004555192
540. GOOG-AT-MDL-004555207
541. GOOG-AT-MDL-006174827
542. GOOG-AT-MDL-006188292
543. GOOG-AT-MDL-006219137
544. GOOG-AT-MDL-006609824
545. GOOG-AT-MDL-006967178
546. GOOG-AT-MDL-007035393
547. GOOG-AT-MDL-007184067
548. GOOG-AT-MDL-007297431
549. GOOG-AT-MDL-007306174
550. GOOG-AT-MDL-007344294
551. GOOG-AT-MDL-007397194
552. GOOG-AT-MDL-007412020
553. GOOG-AT-MDL-008120608
554. GOOG-AT-MDL-008148529
555. GOOG-AT-MDL-008148533
556. GOOG-AT-MDL-008401232
557. GOOG-AT-MDL-008515131
558. GOOG-AT-MDL-008682071
559. GOOG-AT-MDL-008770862
560. GOOG-AT-MDL-008842383
561. GOOG-AT-MDL-008947559
562. GOOG-AT-MDL-008950810
563. GOOG-AT-MDL-008953875
564. GOOG-AT-MDL-008956205
565. GOOG-AT-MDL-008959102
566. GOOG-AT-MDL-008963668
567. GOOG-AT-MDL-008964584
568. GOOG-AT-MDL-008967988
569. GOOG-AT-MDL-008973378
570. GOOG-AT-MDL-008979450
571. GOOG-AT-MDL-008984168
572. GOOG-AT-MDL-008986552
573. GOOG-AT-MDL-008986864
574. GOOG-AT-MDL-008991190
575. GOOG-AT-MDL-008992510
576. GOOG-AT-MDL-009025979
577. GOOG-AT-MDL-009036094
578. GOOG-AT-MDL-009367449
579. GOOG-AT-MDL-009709715
580. GOOG-AT-MDL-012331946
581. GOOG-AT-MDL-012512067
582. GOOG-AT-MDL-012709388
583. GOOG-AT-MDL-012766862
584. GOOG-AT-MDL-012771724
585. GOOG-AT-MDL-012781882
586. GOOG-AT-MDL-013107815
587. GOOG-AT-MDL-013134435
588. GOOG-AT-MDL-013274837
589. GOOG-AT-MDL-013284385
590. GOOG-AT-MDL-013304990
591. GOOG-AT-MDL-013305020
592. GOOG-AT-MDL-014005744
593. GOOG-AT-MDL-014157435
594. GOOG-AT-MDL-014188789
595. GOOG-AT-MDL-014305171
596. GOOG-AT-MDL-014566000
597. GOOG-AT-MDL-014566081
598. GOOG-AT-MDL-014570102
599. GOOG-AT-MDL-014825556
600. GOOG-AT-MDL-015177997
601. GOOG-AT-MDL-015214562
602. GOOG-AT-MDL-015262116
603. GOOG-AT-MDL-015718221
604. GOOG-AT-MDL-015753088
605. GOOG-AT-MDL-015924793
606. GOOG-AT-MDL-015995550
607. GOOG-AT-MDL-017320175
608. GOOG-AT-MDL-017663254
609. GOOG-AT-MDL-017781794
610. GOOG-AT-MDL-018243919
611. GOOG-AT-MDL-018467177
612. GOOG-AT-MDL-018513313
613. GOOG-AT-MDL-018631724
614. GOOG-AT-MDL-018673806
615. GOOG-AT-MDL-018796926
616. GOOG-AT-MDL-019006242
617. GOOG-AT-MDL-019010409
618. GOOG-AT-MDL-019045181
619. GOOG-AT-MDL-019164293
620. GOOG-AT-MDL-019198499
621. GOOG-AT-MDL-019244499
622. GOOG-AT-MDL-019521362
623. GOOG-AT-MDL-019585702
624. GOOG-AT-MDL-019608993
625. GOOG-AT-MDL-019724612
626. GOOG-AT-MDL-B-000744067

627. GOOG-AT-MDL-B-000762975
628. GOOG-AT-MDL-B-001128860
629. GOOG-AT-MDL-B-001605461
630. GOOG-AT-MDL-B-001605538
631. GOOG-AT-MDL-B-001633930
632. GOOG-AT-MDL-B-001831512
633. GOOG-AT-MDL-B-001833845
634. GOOG-AT-MDL-B-001838986
635. GOOG-AT-MDL-B-001856427
636. GOOG-AT-MDL-B-001876817
637. GOOG-AT-MDL-B-002085667
638. GOOG-AT-MDL-B-002090556
639. GOOG-AT-MDL-B-002096327
640. GOOG-AT-MDL-B-002110895
641. GOOG-AT-MDL-B-002235587
642. GOOG-AT-MDL-B-002504803
643. GOOG-AT-MDL-B-002505352
644. GOOG-AT-MDL-B-002508987
645. GOOG-AT-MDL-B-002535093
646. GOOG-AT-MDL-B-002762725
647. GOOG-AT-MDL-B-002763539
648. GOOG-AT-MDL-B-003179191
649. GOOG-AT-MDL-B-003289748
650. GOOG-AT-MDL-B-003741171
651. GOOG-AT-MDL-B-003754823
652. GOOG-AT-MDL-B-004016354
653. GOOG-AT-MDL-B-004246657
654. GOOG-AT-MDL-B-004246668
655. GOOG-AT-MDL-B-004637454
656. GOOG-AT-MDL-B-004864657
657. GOOG-AT-MDL-B-006307109
658. GOOG-AT-MDL-B-006365520
659. GOOG-AT-MDL-B-007244076
660. GOOG-AT-MDL-C-000010785
661. GOOG-DOJ-00300834
662. GOOG-DOJ-13912541
663. GOOG-DOJ-13918991
664. GOOG-DOJ-13931782
665. GOOG-DOJ-13935674
666. GOOG-DOJ-13942082
667. GOOG-DOJ-13949363
668. GOOG-DOJ-13952417
669. GOOG-DOJ-13984617
670. GOOG-DOJ-13990174
671. GOOG-DOJ-13991370
672. GOOG-DOJ-13997420
673. GOOG-DOJ-14039876
674. GOOG-DOJ-14141075
675. GOOG-DOJ-14155094
676. GOOG-DOJ-14156098
677. GOOG-DOJ-14156657
678. GOOG-DOJ-14160868
679. GOOG-DOJ-14231410
680. GOOG-DOJ-14248141
681. GOOG-DOJ-14254613
682. GOOG-DOJ-14265301
683. GOOG-DOJ-14368344
684. GOOG-DOJ-14380896
685. GOOG-DOJ-14398632
686. GOOG-DOJ-14433810
687. GOOG-DOJ-14433907
688. GOOG-DOJ-14435887
689. GOOG-DOJ-14441373
690. GOOG-DOJ-14469279
691. GOOG-DOJ-14473642
692. GOOG-DOJ-14495239
693. GOOG-DOJ-14510835
694. GOOG-DOJ-14510945
695. GOOG-DOJ-14526230
696. GOOG-DOJ-14550176
697. GOOG-DOJ-14551574
698. GOOG-DOJ-14559099
699. GOOG-DOJ-14705058
700. GOOG-DOJ-14706688
701. GOOG-DOJ-14707242
702. GOOG-DOJ-14709267
703. GOOG-DOJ-14712029
704. GOOG-DOJ-14718827
705. GOOG-DOJ-14733010
706. GOOG-DOJ-14733299
707. GOOG-DOJ-14735629
708. GOOG-DOJ-14736531
709. GOOG-DOJ-14743554
710. GOOG-DOJ-14744814
711. GOOG-DOJ-14832749
712. GOOG-DOJ-14879034
713. GOOG-DOJ-14954902
714. GOOG-DOJ-15046502
715. GOOG-DOJ-15064096
716. GOOG-DOJ-15064794
717. GOOG-DOJ-15065751
718. GOOG-DOJ-15071643

719. GOOG-DOJ-15078089
720. GOOG-DOJ-15084805
721. GOOG-DOJ-15085650
722. GOOG-DOJ-15129170
723. GOOG-DOJ-15164364
724. GOOG-DOJ-15359510
725. GOOG-DOJ-15370825
726. GOOG-DOJ-15421777
727. GOOG-DOJ-15421837
728. GOOG-DOJ-15426012
729. GOOG-DOJ-15428402
730. GOOG-DOJ-15433150
731. GOOG-DOJ-15435125
732. GOOG-DOJ-15445619
733. GOOG-DOJ-15446667
734. GOOG-DOJ-15472771
735. GOOG-DOJ-15476271
736. GOOG-DOJ-15603884
737. GOOG-DOJ-15612215
738. GOOG-DOJ-15720586
739. GOOG-DOJ-15769196
740. GOOG-DOJ-15772422
741. GOOG-DOJ-15820805
742. GOOG-DOJ-27571503
743. GOOG-DOJ-27759823
744. GOOG-DOJ-27760500
745. GOOG-DOJ-27762880
746. GOOG-DOJ-28417257
747. GOOG-DOJ-28486313
748. GOOG-DOJ-28501678
749. GOOG-DOJ-29625035
750. GOOG-DOJ-29731245
751. GOOG-DOJ-29803801
752. GOOG-DOJ-32277291
753. GOOG-DOJ-32283742
754. GOOG-DOJ-32311857
755. GOOG-DOJ-32321240
756. GOOG-DOJ-32629220
757. GOOG-DOJ-AT-00337555
758. GOOG-DOJ-AT-00349951
759. GOOG-DOJ-AT-00596359
760. GOOG-DOJ-AT-00597161
761. GOOG-DOJ-AT-00598647
762. GOOG-DOJ-AT-00634965
763. GOOG-DOJ-AT-00916063
764. GOOG-DOJ-AT-01016519
765. GOOG-DOJ-AT-01019307
766. GOOG-DOJ-AT-01140595
767. GOOG-DOJ-AT-01251555
768. GOOG-DOJ-AT-01255154
769. GOOG-DOJ-AT-01288913
770. GOOG-DOJ-AT-01423320
771. GOOG-DOJ-AT-01423537
772. GOOG-DOJ-AT-01481161
773. GOOG-DOJ-AT-01490095
774. GOOG-DOJ-AT-01517573
775. GOOG-DOJ-AT-01526572
776. GOOG-DOJ-AT-01564508
777. GOOG-DOJ-AT-01620514
778. GOOG-DOJ-AT-01682498
779. GOOG-DOJ-AT-01800023
780. GOOG-DOJ-AT-01808791
781. GOOG-DOJ-AT-01809117
782. GOOG-DOJ-AT-01811118
783. GOOG-DOJ-AT-01812188
784. GOOG-DOJ-AT-01827124
785. GOOG-DOJ-AT-01831448
786. GOOG-DOJ-AT-01849832
787. GOOG-DOJ-AT-01892558
788. GOOG-DOJ-AT-01902474
789. GOOG-DOJ-AT-01908449
790. GOOG-DOJ-AT-02137259
791. GOOG-DOJ-AT-02137346
792. GOOG-DOJ-AT-02147591
793. GOOG-DOJ-AT-02149824
794. GOOG-DOJ-AT-02151279
795. GOOG-DOJ-AT-02151860
796. GOOG-DOJ-AT-02156432
797. GOOG-DOJ-AT-02156890
798. GOOG-DOJ-AT-02172942
799. GOOG-DOJ-AT-02190909
800. GOOG-DOJ-AT-02193235
801. GOOG-DOJ-AT-02195580
802. GOOG-DOJ-AT-02195610
803. GOOG-DOJ-AT-02195675
804. GOOG-DOJ-AT-02197386
805. GOOG-DOJ-AT-02199374
806. GOOG-DOJ-AT-02218994
807. GOOG-DOJ-AT-02224601
808. GOOG-DOJ-AT-02225684
809. GOOG-DOJ-AT-02261785
810. GOOG-DOJ-AT-02274670

811. GOOG-DOJ-AT-02323144
812. GOOG-DOJ-AT-02366287
813. GOOG-DOJ-AT-02424328
814. GOOG-DOJ-AT-02467209
815. GOOG-DOJ-AT-02499720
816. GOOG-DOJ-AT-02640072
817. GOOG-NE-02632899
818. GOOG-NE-03597611
819. GOOG-NE-03616755
820. GOOG-NE-03634720
821. GOOG-NE-04384116
822. GOOG-NE-04719370
823. GOOG-NE-05284000
824. GOOG-NE-06580783
825. GOOG-NE-06835928
826. GOOG-NE-07245622
827. GOOG-NE-07249243
828. GOOG-NE-07294175
829. GOOG-NE-09426318
830. GOOG-NE-10291781
831. GOOG-NE-10493534
832. GOOG-NE-11375580
833. GOOG-NE-11849783
834. GOOG-NE-11917922
835. GOOG-NE-12720608
836. GOOG-NE-13197548
837. GOOG-NE-13199159
838. GOOG-NE-13202480
839. GOOG-NE-13203514
840. GOOG-NE-13209197
841. GOOG-NE-13210980
842. GOOG-NE-13212314
843. GOOG-NE-13217070
844. GOOG-NE-13226622
845. GOOG-NE-13231717
846. GOOG-NE-13231861
847. GOOG-NE-13232022
848. GOOG-NE-13351825
849. GOOG-NE-13357217
850. GOOG-NE-13388834
851. GOOG-NE-13394781
852. GOOG-NE-13468541
853. GOOG-NE-13489067
854. GOOG-NE-13499149
855. GOOG-NE-13501797
856. GOOG-NE-13504124
857. GOOG-NE-13505379
858. GOOG-NE-13512082
859. GOOG-NE-13513997
860. GOOG-NE-13603380
861. GOOG-NE-13604113
862. GOOG-NE-13604518
863. GOOG-TEX-00041478
864. GOOG-TEX-00059386
865. GOOG-TEX-00079917
866. GOOG-TEX-00090958
867. GOOG-TEX-00093402
868. GOOG-TEX-00103040
869. GOOG-TEX-00104475
870. GOOG-TEX-00118526
871. GOOG-TEX-00134993
872. GOOG-TEX-00263099
873. GOOG-TEX-00264386
874. GOOG-TEX-00304200
875. GOOG-TEX-00329374
876. GOOG-TEX-00363227
877. GOOG-TEX-00595685
878. GOOG-TEX-00596761
879. GOOG-TEX-00683280
880. GOOG-TEX-00703289
881. GOOG-TEX-00725911
882. GOOG-TEX-00777573
883. GOOG-TEX-00831090
884. GOOG-TEX-00969363
885. GOOG-TEX-00969653
886. GOOG-TEX-01008914
887. GOOG-TEX-01009154
888. GOOG-TEX-01022605
889. GOOG-TEX-01068770
890. GOOG-TEX-01084554
891. GOOG-TEX-01086106
892. GOOG-TEX-01231847
893. GOOG-TX-00001418
894. GOOG-TX-00093439
895. GOOG-TX-00597317
896. [redacted]
897. [redacted]
898. [redacted]
899. [redacted]

Education

- **Ph.D.: Princeton University**. Program in Applied and Computational Mathematics. Title: **Boosting, Margins, and Dynamics**. Advisors: Ingrid Daubechies and Robert Schapire.
- **BS/BA: University at Buffalo (SUNY), Honors Program**. Outstanding Senior Award in the Arts and Sciences (one awarded per year university-wide), separate outstanding senior awards from the Physics Department, Mathematics Department, and Music Department (one awarded per department per year), BS Mathematical Physics, BA Music Theory, Minor in Computer Science, Summa Cum Laude, 1999

Employment History

- **Duke University**, Computer Science Department (50%), Electrical and Computer Engineering Department (50%), Secondary appointments in Statistical Science, Biostatistics & Bioinformatics, and Mathematics. Associate Professor 2016-2019, Professor 2019-present.
- **Massachusetts Institute of Technology**, MIT Computer Science and Artificial Intelligence Laboratory and Sloan School of Management, Associate Professor of Statistics 2013-2016, Assistant Professor of Statistics 2009-2013.
- **Columbia University**, Center for Computational Learning Systems, Associate Research Scientist, 2007-2009.
- **NSF Postdoctoral Research Fellow, New York University**, 2004-2007.

Peer-Reviewed Publications

Papers associated with major awards (winner and finalist)

1. Xuefei Lu, Emanuele Borogonovo, and Cynthia Rudin. **Trustworthy Feature Importance Avoids Unrestricted Permutations**. *Finalist for Data Mining Best Paper Award, INFORMS 2024*, in preparation, 2024.
2. Yiyang Sun, Zhi Chen, Vittorio Orlandi, Tong Wang and Cynthia Rudin. **Sparse and Faithful Explanations without Sparse Models**. *Winner of Data Mining Best Paper Award, INFORMS 2023*, AISTATS, 2024.
3. Edwin Agnew, Michelle Qiu, Lily Zhu, Sam Wiseman and Cynthia Rudin. **The Mechanical Bard: An Interpretable Machine Learning Approach to Shakespearean Sonnet Generation**, Meeting of the Association for Computational Linguistics (ACL), 2023. *Outstanding Paper Award*.
4. Gah-Yi Ban and Cynthia Rudin. *MSOM Best OM paper in OR Award, 2021, INFORMS. (Awarded to the best paper in Operations Research within the last 3 years, awarded by Manufacturing and Service Operations Management Society of INFORMS.)* **The Big Data Newsvendor: Practical Insights from Machine Learning**, Operations Research, Vol. 67, No. 1, pages 90-108, 2019.
5. Aaron F. Struck, Berk Ustun, Andres Rodriguez Ruiz, Jong Woo Lee, Suzette LaRoche, Lawrence J. Hirsch, Emily J Gilmore, Jan Vlachy, Hiba Arif Haider, Cynthia Rudin, M Brandon Westover. *2019 INFORMS Innovative Applications in Analytics Award* (shared with paper below). **Association of an Electroencephalography-Based Risk Score With Seizure Probability in Hospitalized Patients**, JAMA Neurology, 74 (12), 1419-1424, 2017.
6. Berk Ustun and Cynthia Rudin. *2019 INFORMS Innovative Applications in Analytics Award*, also *2017 INFORMS Computing Society Student Paper Prize.* **Learning Optimized Risk Scores**. Journal of Machine Learning Research, 2019. Shorter version **Learning Optimized Risk Scores from Large-Scale Datasets**. Knowledge Discovery in Databases (KDD), 2017.
7. Chaofan Chen, Kangcheng Lin, Cynthia Rudin, Yaron Shaposhnik, Sijia Wang, Tong Wang. **A Holistic Approach to Interpretability in Financial Lending: Models, Visualizations, and Summary-Explanations.** Decision Support Systems, 2021.
   Preliminary work:
   - Chaofan Chen, Kangcheng Lin, Cynthia Rudin, Yaron Shaposhnik, Sijia Wang, Tong Wang. *Winner of the FICO Recognition Award for the Explainable Machine Learning Challenge, 2018.* **An Interpretable Model with Globally Consistent Explanations for Credit Risk.** NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: the Impact of Fairness, Explainability, Accuracy, and Privacy, 2018.

8. Cynthia Rudin and Berk Ustun. *Finalist for 2017 Daniel H. Wagner Prize for Excellence in Operations Research, Institute for Operations Research and Management Science (INFORMS).* **Optimized Scoring Systems: Towards Trust in Machine Learning for Healthcare and Criminal Justice**. INFORMS Journal on Applied Analytics, Special Issue: 2017 Daniel H. Wagner Prize for Excellence in Operations Research Practice, 48(5), pages 399–486, September-October, 2018.
9. William Souillard-Mandar, Randall Davis, Cynthia Rudin, Rhoda Au, David J. Libon, Rodney Swenson, Catherine C. Price, Melissa Lamar, Dana L. Penney. *2016 INFORMS Innovative Applications in Analytics Award* (shared with paper below). **Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test**. Machine Learning, volume 102, number 3, 2016.
10. Berk Ustun and Cynthia Rudin. *2016 INFORMS Innovative Applications in Analytics Award* and also *Runner up, Invenia Labs SEE Award 2018 - Supporting Machine Learning Research with a Positive Impact on Social, Economic, or Environmental (SEE) Challenges*. **Supersparse Linear Integer Models for Optimized Medical Scoring Systems**. Machine Learning, volume 102, number 3, 2016.
11. Cynthia Rudin, Şeyda Ertekin, Rebecca Passonneau, Axinia Radeva, Ashish Tomar, Boyi Xie, Stanley Lewis, Mark Riddle, Debbie Pangsrivinij, Tyler McCormick. *2013 INFORMS Innovative Applications in Analytics Award* **Analytics for Power Grid Distribution Reliability in New York City**. INFORMS Journal on Applied Analytics, volume 44, issue 4, pages 364-383, 2014.
12. Indraneel Mukherjee, Cynthia Rudin, and Robert E. Schapire. *This paper answered an open question published in COLT 2010.* **The Rate of Convergence of AdaBoost**, Journal of Machine Learning Research, volume 14, pages 2315-2347, August 2013.
    Preliminary work:
    - Indraneel Mukherjee, Cynthia Rudin, and Robert E. Schapire. **The Rate of Convergence of AdaBoost**, Proceedings of the 24th Annual Conference on Learning Theory (COLT), 2011.

Publications and other best paper awards (not including those above)

2024

13. Cynthia Rudin, Chudi Zhong, Lesia Semenova, Margo Seltzer, Ronald Parr, Jiachang Liu, Srikar Katta, Jon Donnelly, Harry Chen, Zachery Boner. **Amazing Things Come From Having Many Good Models.** ICML (Spotlight), 2024.
14. Alina Jade Barnett, Zhicheng Guo, Jin Jing, Wendong Ge, Cynthia Rudin, M. Brandon Westover. **Improving Clinician Performance in Classification of EEG Patterns on the Ictal-Interictal-Injury Continuum using Interpretable Machine Learning**, New England Journal of Medicine AI (NEJM AI), 2024.
15. Jon Donnelly, Jon, Luke Moffett, Alina Jade Barnett, Hari Trivedi, Fides Regina Schwartz, Joseph Lo, Cynthia Rudin. **AsymMirai: Interpretable Mammography-Based Deep Learning Model for 1- to 5-year Breast Cancer Risk Prediction**, Radiology, 2024.
16. Sully F. Chen, Zhicheng Guo, Cheng Ding, Xiao Hu, Cynthia Rudin. **Learned Kernels for Sparse, Interpretable, and Efficient Medical Time Series Processing**. Nature Machine Intelligence, 2024.
17. Lesia Semenova, Yingfan Wang, Shane Falcinelli, Nancie Archin, Alicia D Cooper-Volkheimer, David M Margolis, Nilu Goonetilleke, David M Murdoch, Cynthia D Rudin, and Edward P Browne. **Machine Learning Approaches Identify Immunologic Signatures of Total and Intact HIV DNA During Long-Term Antiretroviral Therapy**, eLife, 2024.
18. Travis Seale-Carlisle, Saksham Jain, Courtney Lee, Caroline Levenson, Swathi Ramprasad, Brandon Garrett, Sudeepa Roy, Cynthia Rudin, Alexander Volfovsky **Evaluating Pre-trial Programs Using Machine Learning Matching Algorithms**, AAAI (Oral), 2024.
19. Siong Thye Goh, Lesia Semenova, and Cynthia Rudin. **Sparse Density Trees and Lists: An Interpretable Alternative to High-Dimensional Histograms**, INFORMS Journal on Data Science, 2024.
20. Srikar Katta, Harsh Parikh, Cynthia Rudin, Alexander Volfovsky. **Interpretable Causal Inference for Analyzing Wearable, Sensor, and Distributional Data**. *2024 Joint Statistical Meeting Paper Award, American Statistical Association, Biometrics Section*, AISTATS, 2024.

21. Rui Zhang, Rui Xin, Margo Seltzer, and Cynthia Rudin. **Optimal Sparse Survival Trees**, AISTATS, 2024.
22. Harsh Parikh, Quinn Lanners, Zade Akras, Sahar Zafar, M Brandon Westover, Cynthia Rudin, Alexander Volfovsky, **Safe and Interpretable Estimation of Optimal Treatment Regimes**, AISTATS, 2024.
23. Stephen Hahn, Jerry Yin, Rico Zhu, Weihan Xu, Yue Jiang, Simon Mak, Cynthia Rudin. **SentHYMNent: An Interpretable and Sentiment-Driven Model for Algorithmic Melody Harmonization**, KDD, 2024.
24. Brandon L. Garrett and Cynthia Rudin. **The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice**, Cornell Law Review, 2024.
25. Cheng Ding, Zhicheng Guo, Cynthia Rudin, Ran Xiao, Amit Shah, Duc H. Do, Randall J Lee, Gari Clifford, Fadi B Nahab, Xiao Hu. **Learning From Alarms: A Robust Learning Approach for Accurate Photoplethysmography-Based Atrial Fibrillation Detection using Eight Million Samples Labeled with Imprecise Arrhythmia Alarms**, IEEE Journal of Biomedical and Health Informatics (JBHI), 2024.
26. Pranay Jain, Cheng Ding, Cynthia Rudin, and Xiao Hu. **A Self-Supervised Algorithm for Denoising Photoplethysmography Signals for Heart Rate Estimation from Wearables**, Harvard Data Science Review, 2024.
27. Ashokkumar Manickam, Jackson J Peterson, Yuriko Harigaya, David M Murdoch, David M Margolis, Alex Oesterling, Zhicheng Guo, Cynthia D Rudin, Yuchao Jiang, and Edward P Browne. **Integrated single-cell multiomic analysis of HIV latency reversal reveals novel regulators of viral reactivation**, Genomics, Proteomics, and Bioinformatics, 2024.
28. Neha R. Gupta, Vittorio Orlandi, Chia-Rui Chang, Tianyu Wang, Marco Morucci, Pritam Dey, Thomas J. Howell, Xian Sun, Angikar Ghosal, Sudeepa Roy, Cynthia Rudin, Alexander Volfovsky. **dame-flame: A Python Library Providing Fast Interpretable Matching for Causal Inference.** Journal of Statistical Software, accepted, 2024.
    - Software for FLAME by our students Vittorio Orlandi and Neha Gupta was the *Honorable Mention for the 2022 John M. Chambers Statistical Software Award from the American Statistical Association.*
29. Harsh Parikh, Haoqi Sun, Rajesh Amerineni, Eric S. Rosenthal, Alexander Volfovsky, Cynthia Rudin, M. Brandon Westover, Sahar F. Zafar. **How Many Patients Do You Need? Investigating Trial Designs for Anti-Seizure Treatment in Acute Brain Injury Patients**, Annals of Clinical and Translational Neurology, 2024.
30. Julia Yang, Alina Barnett, Jonathan Donnelly, Satvik Kishore, Jerry Fang, Fides Schwartz, Chaofan Chen, Joseph Lo, Cynthia Rudin. **FPN-IAIA-BL: A Multi-Scale Interpretable Deep Learning Model for Classification of Mass Margins in Digital Mammography** CVPR 2024 Workshop on Domain adaptation, Explainability, Fairness in AI for Medical Image Analysis (DEF-AI-MIA), oral presentation, 2024.
31. Ron Parr, Michal Moshkovitz, Lesia Semenova, Harry Chen, Zack Boner, Cynthia Rudin. **Transition Noise Facilitates Interpretability**, Workshop on Interpretable Policies in Reinforcement Learning (InterpPol) @ RLC-2024 (Oral).
32. Stephen Hahn, Weihan Xu, Zirui Yin, Rico Zhu, Simon Mak, Yue Jiang, Cynthia Rudin. **A New Dataset, Notation Software, and Representation for Computational Schenkerian Analysis**. International Society for Music Information Retrieval (ISMIR) Conference, 2024.

2023

33. Brandon L. Garrett and Cynthia Rudin. **Interpretable Algorithmic Forensics**, The Proceedings of the National Academy of Sciences (PNAS), 2023.
34. Lesia Semenova, Harry Chen, Ronald Parr, Cynthia Rudin. **A Path to Simpler Models Starts With Noise**, NeurIPS, 2023.
35. Zhi Chen, Chudi Zhong, Margo Seltzer, Cynthia Rudin. **Exploring and Interacting with the Set of Good Sparse Generalized Additive Models**, NeurIPS, 2023.
36. Chiyu Ma, Brandon Zhao, Chaofan Chen, Cynthia Rudin. **This Looks Like Those: Illuminating Prototypical Concepts Using Multiple Visualizations**, NeurIPS, 2023.
37. Jiachang Liu, Sam Rosen, Chudi Zhong, Cynthia Rudin. **OKRidge: Scalable Optimal k-Sparse Ridge Regression for Learning Dynamical Systems**, NeurIPS (Spotlight), 2023.

- 2024 INFORMS Computing Society Student Paper Award Runner Up, 2024

38. Jon Donnelly, Srikar Katta, Cynthia Rudin, Edward P Browne. **The Rashomon Importance Distribution: Getting RID of Unstable, Single Model-based Variable Importance**, NeurIPS (Spotlight), 2023.
39. Stephen Hahn, Rico Zhu, Yue Jiang, Simon Mak, and Cynthia Rudin. **New Orleans: An Adventure in Music**, NeurIPS Creative AI track. (demo) 2023.
40. Stephen Hahn, Rico Zhu, Simon Mak, Cynthia Rudin, Yue Jiang. **An Interpretable, Flexible, and Interactive Probabilistic Framework for Melody Generation**, KDD 2023.
41. Dennis Tang, Frank Willard, Ronan Tegerdine, Luke Triplett, Jon Donnelly, Luke Moffett, Lesia Semenova, Alina Jade Barnett, Jin Jing, Cynthia Rudin, and M. Brandon Westover. **ProtoEEGNet: An Interpretable Approach for Detecting Interictal Epileptiform Discharges**, Medical Imaging meets NeurIPS Workshop (MedNeurIPS), (Oral), 2023.
42. Samantha M. McDonald, Emily K. Augustine, Quinn Lanners, Cynthia Rudin, L. Catherine Brinson, and Matthew L. Becker. **Applied Machine Learning as a Driver for Polymeric Biomaterials Design**, Nature Communications, 2023.
43. Jacob Peloquin, Alina Kirillova, Cynthia Rudin, L.C. Brinson, Ken Gall. **Prediction of tensile performance for 3D printed photopolymer gyroid lattices using structural porosity, base material properties, and machine learning**, Materials & Design, 2023.
44. Jacob Peloquin, Alina Kirillova, Elizabeth Mathey, Cynthia Rudin, L. Catherine Brinson, Ken Gall. **Tensile performance data of 3D printed photopolymer gyroid lattices**, Data in Brief, 2023.
45. Quinn Lanners, Harsh Parikh, Alexander Volfovsky, Cynthia Rudin, David Page. **Variable Importance Matching for Causal Inference**, UAI, 2023.
    A shorter version appears at this workshop:
    - Quinn Lanners, Harsh Parikh, Alexander Volfovsky, Cynthia Rudin, and David Page. **Matching Using Feature Importance: An Auditable Approach to Causal Inference**, 9th International Conference on Computational Social Science IC^2S^2, 2023
46. Harsh Parikh, Kentaro Hoffman, Haoqi Sun, Wendong Ge, Jin Jing, Rajesh Amerineni, Lin Liu, Jimeng Sun, Sahar Zafar, Aaron Struck, Alexander Volfovsky, Cynthia Rudin, M. Brandon Westover. **Effects of Epileptiform Activity on Discharge Outcome in Critically Ill Patients: A Retrospective Cross-Sectional Study**, The Lancet Digital Health, 2023.
47. Rui Zhang, Rui Xin, Margo Seltzer, and Cynthia Rudin. **Optimal Sparse Regression Trees**, AAAI, 2023.
48. Cynthia Rudin and Yaron Shaposhnik. **Globally-Consistent Rule-Based Summary-Explanations for Machine Learning Models: Application to Credit-Risk Evaluation**, Journal of Machine Learning Research, 2023.
    - An earlier version appeared at the Conference on Information Systems and Technology (CIST), 2019.
49. Fernanda Bravo, Cynthia Rudin, Yuting Yuan and Yaron Shaposhnik. **Interpretable Prediction Rules for Congestion Risk in ICUs**. Stochastic Systems, 2023.
    - Conference version: Fernanda Bravo, Cynthia Rudin, Yuting Yuan and Yaron Shaposhnik. **Simple Rules for Predicting Congestion Risk in Queueing Systems: Application to ICUs**, 2019 INFORMS Workshop on Data Science (DS 2019), (Oral).
50. Shane D Falcinelli, Alicia Volkheimer, Lesia Semenova, Ethan Wu, Alexander Richardson, Manickam Ashokkumar, David M Margolis, Nancie M Archin, Cynthia D Rudin, David Murdoch, Edward P Browne. **Impact of cannabis use on immune cell populations and the viral reservoir in people with HIV on suppressive antiretroviral therapy**, The Journal of Infectious Disease (JID), 2023.
51. Yanchen Jessie Ou, Alina Jade Barnett, Anika Mitra, Fides Regina Schwartz, Chaofan Chen, Lars Grimm, Joseph Y. Lo, Cynthia Rudin. **A user interface to communicate interpretable AI decisions to radiologists**. Proceedings of SPIE Medical Imaging, 2023.

52. Edward W. Felten, Jennifer Mnookin, Thomas D. Albright, Ricardo Baeza-Yates, Bob Blakey, Patrick Grother, Marvin B. Haiman, Aziz Z. Huq, Anil K. Jain, Elizabeth E. Joh, Michael C. King, Nicol Turner Lee, Ira S. Reese, and Cynthia Rudin. **Facial Recognition Technology: Current Capabilities, Future Prospects, and Governance,** National Academies Press, 2024. (Consensus study report of the National Academies of Sciences Engineering, and Medicine)

2022

53. Rui Xin, Chudi Zhong, Zhi Chen, Takuya Takagi, Margo Seltzer, Cynthia Rudin. **Exploring the Whole Rashomon Set of Sparse Decision Trees**, *Finalist for INFORMS 2022 Data Mining Best Paper Competition Award, Student Track*, NeurIPS (Oral), 2022.
54. Jiachang Liu, Chudi Zhong, Boxuan Li, Margo Seltzer, Cynthia Rudin. **FasterRisk: Fast and Accurate Interpretable Risk Scores**, NeurIPS, 2022.
55. Zijie Wang, Chudi Zhong, Rui Xin, Takuya Takagi, Zhi Chen, Duen Horng Chau, Cynthia Rudin and Margo Seltzer. **TimberTrek: Exploring and Curating Trustworthy Decision Trees with Interactive Visualization**, IEEE VIS: Visualization & Visual Analytics, 2022.
56. Harsh Parikh, Cynthia Rudin, Alexander Volfovsky. **MALTS: Matching After Learning to Stretch.** Journal of Machine Learning Research, 2022
57. Haiyang Huang, Yingfan Wang, Cynthia Rudin, and Edward Browne. **Towards a Comprehensive Evaluation of Dimension Reduction Methods for Transcriptomic Data Visualization.** Communications Biology (Nature), 2022.
58. Vaishali Jain, Ted Enamorado, and Cynthia Rudin. **The Importance of being Ernest, Ekundayo, or Eswari: An Interpretable Machine Learning Approach to Name-based Ethnicity Classification**, Harvard Data Science Review, 2022.
59. Elita Lobo, Harvineet Singh, Marek Petrik, Cynthia Rudin, Himabindu Lakkaraju. **Data Poisoning Attacks on Off-Policy Policy Evaluation Methods**, (Oral - top 5%), UAI, 2022.
60. Lesia Semenova, Cynthia Rudin, and Ron Parr. **On the Existence of Simpler Machine Learning Models.** ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT), 2022.
61. Cynthia Rudin. **Why Black Box Machine Learning Should be Avoided for High Stakes Decisions, in Brief**, Nature Reviews Methods Primers, 2022.
62. Jiachang Liu, Chudi Zhong, Margo Seltzer, and Cynthia Rudin. **Fast Sparse Classification for Generalized Linear and Additive Models**, AISTATS, 2022.
63. Cynthia Rudin, Chaofan Chen, Zhi Chen, Haiyang Huang, Lesia Semenova, Chudi Zhong. **Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges**, Statistics Surveys, 2022.
64. Hayden McTavish, Chudi Zhong, Reto Achermann, Ilias Karimalis, Jacques Chen, Cynthia Rudin, Margo Seltzer. **Fast Sparse Decision Tree Optimization via Reference Ensembles**, AAAI, 2022.
65. Zhi Chen, Alexander Ogren, Chiara Daraio, L. Catherine Brinson, Cynthia Rudin. *Winner of the 2022 Physical and Engineering Sciences (SPES) and the Quality and Productivity (Q&P) Student Paper Competition of the American Statistical Association.* **How to See Hidden Patterns in Metamaterials with Interpretable Machine Learning**. Extreme Mechanics Letters, 2022.
66. Tong Wang and Cynthia Rudin. **Subgroup Identification for Enhanced Treatment Effect with Decision Rules**, INFORMS Journal on Computing, 2022.
67. Caroline Wang, Bin Han, Bhrij Patel, Feroze Mohideen, Cynthia Rudin. **In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction**. Journal of Quantitative Criminology, 2022.
68. Masoud Afnan, Michael Anis Mihdi Afnan, Yanhe Liu, Julian Savulescu, Abhishek Mishra, Vincent Conitzer, Cynthia Rudin. **Data solidarity for machine learning for embryo selection; a call for the creation of an open access repository of embryo data**. Reproductive BioMedicine Online, 2022.
69. Chunxiao Li, Cynthia Rudin, and Tyler McCormick. **Rethinking Nonlinear Instrumental Variable Models through Prediction Validity**, Journal of Machine Learning Research, 2022.

70. Marco Morucci, Md. Noor-E-Alam, and Cynthia Rudin. **A Robust Approach to Quantifying Uncertainty in Matching Problems of Causal Inference**, INFORMS Journal on Data Science, 2022.
71. Edwin Agnew, Lily Zhu, Sam Wiseman, and Cynthia Rudin. **Can a Computer Really Write Poetry?** Harvard Data Science Review, 2022.
72. Ali Behrouz, Mathias Lécuyer, Cynthia Rudin and Margo Seltzer. **Fast Optimization of Weighted Sparse Decision Trees for use in Optimal Treatment Regimes and Optimal Policy Design**, Workshop on Advances in Interpretable Machine Learning and Artificial Intelligence (AIMLAI) at CIKM, 2022.
73. Gaurav Rajesh Parikh, Jenny Huang, Albert Sun, Lesia Semenova, Cynthia Rudin. *Winner of the American Statistical Association Data Challenge Expo (Student Winners), 2022,* **Making the World More Equal, One Ride at a Time: Studying Public Transportation Initiatives Using Interpretable Causal Inference.** NeurIPS 2022 Workshop on Causality for Real-world Impact (CML4Impact), 2022.

2021

74. Cynthia Rudin and Shawn Bushway. **A Truth Serum for your Personal Perspective on Facial Recognition Software in Law Enforcement**, Translational Criminology, Fall 2021
75. Tianyu Wang and Cynthia Rudin. *Best Paper Second Prize for INFORMS Data Mining Best Paper Competition - General Track 2021*, **Bandit Learning for Proportionally Fair Allocations**, 2021.
76. Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. **Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP for Data Visualization**, Journal of Machine Learning Research, 2021.
    - Our PaCMAP software is the winner of the *2023 John M. Chambers Statistical Software Award* from the American Statistical Association and the *2024 Award for Innovation in Statistical Programming and Analytics (AISPAA)* from the American Statistical Association.
77. Tianyu Wang, Marco Morucci, M. Usaid Awan, Yameng Liu, Sudeepa Roy, Cynthia Rudin, Alexander Volfovsky. **FLAME: A Fast Large-scale Almost Matching Exactly Approach to Causal Inference**. Journal of Machine Learning Research, 2021.
    - Software for FLAME by our students Vittorio Orlandi and Neha Gupta was the *Honorable Mention for the 2022 John M. Chambers Statistical Software Award* from the American Statistical Association.
78. Stefano Tracà, Cynthia Rudin, Weiyu Yan. *Best paper award, INFORMS 2016 Data Mining & Decision Analytics Workshop,* also *Finalist for 2015 IBM Service Science Best Paper Award.* **Regulating Greed over Time in Multi-Armed Bandits.** Journal of Machine Learning Research, 2021.
79. Jianyou Wang, Xiaoxuan Zhang, Yuren Zhou, Christopher Suh, Cynthia Rudin. **There Once Was a Really Bad Poet, It Was Automated But You Didn't Know It.** Transactions of the Association for Computational Linguistics (TACL), 2021. (Presented at ACL, 2021)
80. Divya Koyyalagunta, Anna Sun, Rachel Lea Draelos, Cynthia Rudin. **Playing Codenames with Language Graphs and Word Embeddings.** Journal of Artificial Intelligence Research (JAIR), 2021.
81. Michael Anis Mihdi Afnan, Cynthia Rudin, Vincent Conitzer, Julian Savulescu, Abhishek Mishra, Yanhe Liu and Masoud Afnan. **Ethical Implementation of Artificial Intelligence to Select Embryos in In Vitro Fertilization.** Fourth AAAI/ACM Conference on AI, Ethics, and Society (AIES), 2021. (Our related work "Embryo selection using Artificial Intelligence (AI): Epistemic and ethical considerations" accepted for oral presentation at European Society of Human Reproduction and Embryology, ESHRE, 2021, and our related work "Embryo selection by "black-box" artificial intelligence: The ethical and epistemic considerations" accepted for oral presentation at Fertility Society of Australia and New Zealand Annual Conference, 2021)
82. Michael Anis Mihdi Afnan, Yanhe Liu, Vincent Conitzer, Cynthia Rudin, Abhishek Mishra, Julian Savulescu, Masoud Afnan. **Interpretable, Not Black-Box, Artificial Intelligence Should be Used for Embryo Selection**, Human Reproduction Open, 2021.

83. Alina Jade Barnett, Fides Regina Schwartz, Chaofan Tao, Chaofan Chen, Yinhao Ren, Joseph Y. Lo, Cynthia Rudin. **IAIA-BL: A Case-based Interpretable Deep Learning Model for Classification of Mass Lesions in Digital Mammography**, Nature Machine Intelligence, 2021. Conference papers related to this project:
    - Alina Jade Barnett, Fides Regina Schwartz, Chaofan Tao, Chaofan Chen, Yinhao Ren, Joseph Y. Lo, and Cynthia Rudin. **Interpretable Mammographic Image Classification using Cased-Based Reasoning and Deep Learning**, Oral Presentation, IJCAI-21 Workshop on Deep Learning, Case-Based Reasoning, and AutoML: Present and Future Synergies, 2021.
    - Alina Jade Barnett, Vaibhav Sharma, Neel Gajjar, Jerry Fang, Fides Schwartz M.D., Chaofan Chen, Joseph Y. Lo, Cynthia Rudin. **A user interface to communicate interpretable AI decisions to radiologists**. SPIE Medical Imaging, 2022.
84. Zhicheng Guo, Cheng Ding, Xiao Hu, and Cynthia Rudin. **A Supervised Machine Learning Semantic Segmentation Approach for Detecting Artifacts in Plethysmography Signals from Wearables**. Physiological Measurement, 2021.
85. Alex Oesterling, Angikar Ghosal, Haoyang Yu, Rui Xin, Yasa Baig, Lesia Semenova, and Cynthia Rudin. **Duke Data Science 3C Citation Classification Task B**, (Outlines our strategy for our *third-place competition entry in the 2021 3C Shared Task Citation Context Classification based on Purpose competition*), Oral Presentation, Second Workshop on Scholarly Document Processing (SDP) Workshop, NAACL, 2021

2020

86. Zhi Chen, Yijie Bei and Cynthia Rudin. **Concept Whitening for Interpretable Image Recognition.** Nature Machine Intelligence, 2020.
87. Jiayun Dong and Cynthia Rudin. **Variable Importance Clouds: A Way to Explore Variable Importance for the Set of Good Models**, Nature Machine Intelligence, 2020
88. Beau Coker, Cynthia Rudin and Gary King. **A Theory of Statistical Inference for Ensuring the Robustness of Scientific Results.** Management Science, 2020.
89. Jimmy Lin, Chudi Zhong, Diane Hu, Cynthia Rudin, Margo Seltzer. **Generalized and Scalable Optimal Sparse Decision Trees.** ICML, 2020.
90. Tianyu Wang and Cynthia Rudin. **Bandits for BMO Functions.** ICML, 2020.
91. Marco Morucci, Vittorio Orlandi, Sudeepa Roy, Cynthia Rudin, Alexander Volfovsky. **Adaptive Hyper-box Matching for Interpretable Individualized Treatment Effect Estimation.** UAI, 2020.
92. Cynthia Rudin, Caroline Wang and Beau Coker. **The *Age* of Secrecy and Unfairness in Recidivism Prediction.** Harvard Data Science Review, 2020.
    - Broader Issues Surrounding Model Transparency in Criminal Justice Risk Scoring. (Rejoinder) Harvard Data Science Review, 2020.
93. Sachit Menon, Alexandru Damian, Nikhil Ravi, Shijia Hu, Cynthia Rudin. **PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models**. CVPR, 2020.
94. Marco Morucci, Vittorio Orlandi, Sudeepa Roy, Alexander Volfovsky. **Almost-Matching-Exactly for Treatment Effect Estimation under Network Interference**. AISTATS, 2020.
95. Tianyu Wang, Weicheng Ye, Dawei Geng and Cynthia Rudin. **Towards Practical Lipschitz Bandits**. Foundations of Data Science (FODS), 2020.
96. Jerry Liu, Nathan O'Hara, Alexander Rubin, Rachel Draelos, Cynthia Rudin. **Metaphor Detection Using Contextual Word Embeddings from Transformers**. Proceedings of the Second Workshop on Figurative Language Processing, pages 250-255, 2020.
97. Hunter Gregory, Steven Li, Pouya Mohammadi, Natalie Tarn, Rachel Draelos, Cynthia Rudin. **A Transformer Approach to Contextual Sarcasm Detection in Twitter.** Proceedings of the Second Workshop on Figurative Language Processing, pages 270-275, 2020.

98. Qinwen Huang, Ye Zhou, Xiaochen Du, Reed Chen, Jianyou Wang, Cynthia Rudin, Alberto Bartesaghi. **Cryo-ZSSR: Multiple-Image Super-Resolution Based on Deep Internal Learning**. Machine Learning for Structural Biology Workshop at the 34th Conference on Neural Information Processing Systems, 2020.

2019

99. Cynthia Rudin and Joanna Radin. **Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition**. Harvard Data Science Review, 2019.
100. Aaron Fisher, Cynthia Rudin, Francesca Dominici. **All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously**. Journal of Machine Learning Research, 2019.
101. Harsh Parikh, Cynthia Rudin, and Alexander Volfovsky. **An Application of Matching After Learning To Stretch (MALTS) to the ACIC 2018 Causal Inference Challenge Data**. Observational Studies, Issue 5, pages 118-130, 2019.
102. Chaofan Chen, Oscar Li, Chaofan Tao, Alina Jade Barnett, Jonathan Su, Cynthia Rudin. **This Looks Like That: Deep Learning for Interpretable Image Recognition**. NeurIPS (Spotlight), 2019.
103. Xiyang Hu, Cynthia Rudin, and Margo Seltzer. **Optimal Sparse Decision Trees**. NeurIPS (Spotlight), 2019.
104. Peter Hase, Chaofan Chen, Oscar Li, Cynthia Rudin. **Interpretable Image Recognition with Hierarchical Prototypes**. AAAI Human Computation (AAAI-HCOMP), 2019.
105. Cynthia Rudin. **Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.** Nature Machine Intelligence, 2019.
    - Shorter version called **Please Stop Explaining Black Box Machine Learning Models for High Stakes Decisions** appeared at NIPS 2018 Workshop on Critiquing and Correcting Trends in Machine Learning, 2018.
106. Stefano Tracà, Weiyu Yan, and Cynthia Rudin. **Reducing Exploration of Dying Arms in Mortal Bandits**. UAI, 2019
107. Usaid Awan, Yameng Liu, Marco Morucci, Sudeepa Roy, Cynthia Rudin, and Alexander Volfovsky. **Interpretable Almost-Exact Matching With Instrumental Variables.** UAI, 2019
108. Awa Dieng, Yameng Liu, Sudeepa Roy, Cynthia Rudin, and Alexander Volfovsky. **Interpretable Almost-Exact Matching For Causal Inference.** AISTATS, 2019.

2018

109. Fulton Wang, Tyler McCormick, Cynthia Rudin, and John Gore. *Best Poster Award, Conference of the ASA Section on Statistical Learning and Data Mining, 2014.* **Modeling Recovery Curves With Application to Prostatectomy.** Biostatistics, 2018.
110. Cynthia Rudin and Yining Wang. *Finalist for 2017 QSR (Quality, Reliability and Statistics) best refereed paper competition, INFORMS 2017.* **On Direct Learning to Rank and Rerank**. Artificial Intelligence and Statistics (AISTATS), 2018.
111. John Benhardt, Tianlin Duan, Peter Hase, Liuyi Zhu, Cynthia Rudin. *Winner of the 2018 PoetiX Literary Turing Test Award for computer-generated poetry.* **Shall I Compare Thee to a Machine-Written Sonnet? An Approach to Algorithmic Sonnet Generation,** 2018.
112. Yijie Bei, Alex Damian, Shijia Hu, Sachit Menon, Nikhil Ravi, and Cynthia Rudin. *NTIRE-CVPR 2018 Image Super-Resolution Challenge: winner for Track 1 (classic bicubic), honorable mention for Track 2 (realistic mild adverse conditions)*. **New Techniques for Preserving Global Structure and Denoising with Low Information Loss in Single-Image Super-Resolution**, New Trends in Image Restoration and Enhancement Workshop and Challenges on Super-Resolution, Dehazing, and Spectral Reconstruction, NTIRE-CVPR, 2018.
113. Cynthia Rudin and Şeyda Ertekin. **Learning Customized and Optimized Lists of Rules with Mathematical Programming**. Mathematical Programming C (Computation), Mathematical Programming Computation, Volume 10, Number 4, pages 659-702, 2018.

114. Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin. **Learning Certifiably Optimal Rule Lists for Categorical Data**, Journal of Machine Learning Research, volume 18, no. 234, pages 1-78, 2018.

115. Nicholas Larus-Stone, Elaine Angelino, Daniel Alabi, Margo Seltzer, Vassilios Kaxiras, Aditya Saligrama and Cynthia Rudin. **Systems Optimizations for Learning Certifiably Optimal Rule Lists**. Conference on Machine Learning and Systems (MLSys), 2018.

116. Oscar Li, Hao Liu, Chaofan Chen, and Cynthia Rudin. **Deep Learning for Case-based Reasoning through Prototypes: A Neural Network that Explains its Predictions**. Association for the Advancement of Artificial Intelligence (AAAI), 2018.

117. Chaofan Chen and Cynthia Rudin. **An Optimization Approach to Learning Falling Rule Lists**. Artificial Intelligence and Statistics (AISTATS), 2018.

2017

118. Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin. **Certifiably Optimal Rule Lists for Categorical Data**, Knowledge Discovery in Databases (KDD - oral presentation), 2017.

119. Berk Ustun, Lenard A. Adler, Cynthia Rudin, Stephen V. Faraone, Thomas J. Spencer, Patricia Berglund, Michael J. Gruber, Ronald C. Kessler. **The World Health Organization Adult Attention-Deficit/Hyperactivity Disorder Self-Report Screening Scale for DSM-5**. JAMA Psychiatry, April 2017.

120. Tong Wang, Cynthia Rudin, Finale Doshi, Yimin Liu, Erica Klampfl, and Perry MacNeille. **Bayesian Rule Sets for Interpretable Classification, with Application to Context-Aware Recommender Systems**. Journal of Machine Learning Research (JMLR), volume 18, number 70, pages 1-37, 2017.

121. Fulton Wang and Cynthia Rudin. **Causal Falling Rule Lists**, Fairness, Accountability, and Transparency (FATML), 2017 (longer version on ArXiv).

122. Hongyu Yang, Cynthia Rudin, and Margo Seltzer. *Winner of Student Paper Competition sponsored by the Statistical Learning and Data Mining section (SLDM) of the American Statistical Association, 2016.* **Scalable Bayesian Rule Lists**. International Conference on Machine Learning (ICML), 2017.

123. Himabindu Lakkaraju and Cynthia Rudin. *Finalist for 2017 INFORMS Data Mining Best Paper Competition.* **Learning Cost-Effective and Interpretable Treatment Regimes**. Artificial Intelligence and Statistics (AISTATS), 2017.
    - Shorter versions accepted to workshops: NIPS Machine Learning for Healthcare (ML4HC), NIPS Workshop on Interpretable Machine Learning in Complex Systems, and NIPS Workshop on ML and the Law, NIPS 2016.

2016

124. Jiaming Zeng, Berk Ustun, and Cynthia Rudin. *Winner of 2015 Undergraduate Statistics Research Project Competition (USRESP) sponsored by the American Statistical Association (ASA) and the Consortium for Advancement of Undergraduate Statistics Education (CAUSE).* **Interpretable Classification Models for Recidivism Prediction.** Journal of the Royal Statistical Society Series A, volume 180, issue 3, 2016.

125. Tong Wang, Cynthia Rudin, Finale Doshi, Yimin Liu, Erica Klampfl, and Perry MacNeille. **Bayesian Rule Sets for Interpretable Classification**. IEEE International Conference on Data Mining (ICDM), 2016.

126. Ramin Moghaddass, Cynthia Rudin, and David Madigan. **The Factorized Self-Controlled Case Series Method: An Approach for Estimating the Effects of Many Drugs on Many Outcomes.** Journal of Machine Learning Research, 17(185):1-24, 2016.

127. Berk Ustun, M. Brandon Westover, Cynthia Rudin, and Matt T. Bianchi. **Clinical Prediction Models for Sleep Apnea: The Importance of Medical History over Symptoms**. Journal of Clinical Sleep Medicine, volume 12, number 2, 2016.

128. Vikas Garg, Cynthia Rudin, and Tommi Jaakola. **CRAFT: ClusteR-specific Assorted Feature selecTion**, Artificial Intelligence and Statistics (AISTATS), 2016.

129. Benjamin Letham, Portia A. Letham, Cynthia Rudin, Edward P. Browne. **Prediction Uncertainty and Optimal Experimental Design for Learning Dynamical Systems.** Chaos, volume 26, number 6, 2016.

130. Benjamin Letham, Lydia M. Letham and Cynthia Rudin. **Bayesian Inference of Arrival Rate and Substitution Behavior from Sales Transaction Data with Stockouts**. Knowledge Discovery in Databases (KDD), 2016
131. Edward P. Browne, Benjamin Letham, and Cynthia Rudin. **A Computational Model of Inhibition of HIV-1 by Interferon-Alpha**. PLoS ONE, vol 11, no 3, pages 1–16, March, 2016.
132. Şeyda Ertekin and Cynthia Rudin. **A Bayesian Approach to Learning Scoring Systems**. Big Data, volume 3, number 4, 2016.
133. John Guckenheimer, Thomas Overbye (Co-chairs), and committee: Daniel Bienstock, Anjan Bose, Terry Boston, Jeffery Dagle, Marija D. Ilic, Christopher K. Jones, Frank P. Kelly, Yannis G. Kevrekidis, Ralph D. Masiello, Juan C. Meza, Cynthia Rudin, Robert J. Thomas, and Margaret H. Wright. **Analytic Research Foundations for the Next-Generation Electric Grid.** The National Academies Press, 2016. (Consensus study report of the National Academies of Sciences Engineering, and Medicine)

2015

134. Fulton Wang and Cynthia Rudin. *Winner of Best Student Paper Competition, Statistical Learning and Data Mining section (SLDM) of the American Statistical Association, 2015,* also *Finalist for INFORMS Data Mining Section Best Student Paper Award, 2015.* **Falling Rule Lists**. Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), 2015.
135. Benjamin Letham, Cynthia Rudin, Tyler H. McCormick, and David Madigan. *Winner of Data Mining Best Student Paper Competition, INFORMS 2013,* also *Winner of Student Paper Competition sponsored by the Statistical Learning and Data Mining section (SLDM) of the American Statistical Association, 2014.* **Building Interpretable Classifiers with Rules using Bayesian Analysis: Building a Better Stroke Prediction Model**. Annals of Applied Statistics, volume 9, number 3, pages 1350-1371, 2015.
Shorter version:
    - Benjamin Letham, Cynthia Rudin, Tyler McCormick, and David Madigan. **An Interpretable Model for Stroke Prediction Using Rules and Bayesian Analysis.** Proceedings of 2014 KDD Workshop on Data Science for Social Good, 2014
136. Tong Wang, Cynthia Rudin, Daniel Wagner and Rich Sevieri. *Second place in INFORMS 2015 Doing Good with Good OR Paper Competition.* **Finding Patterns with a Rotten Core: Data Mining for Crime Series with Core Sets**. Big Data. volume 3, issue 1, 2015. (Special issue on Data for Social Good)
137. Ramin Moghaddass and Cynthia Rudin. **The Latent State Hazard Model, with Application to Wind Turbine Reliability**. Annals of Applied Statistics, volume 9, number 4, pages 1823–1863, 2015.
138. Şeyda Ertekin, Cynthia Rudin, and Tyler McCormick. **Reactive Point Processes: A New Approach to Predicting Power Failures in Underground Electrical Systems**. Annals of Applied Statistics, volume 9, No 1, pages 122–144, 2015.

2014

139. Theja Tulabandhula and Cynthia Rudin. **Tire Changes, Fresh Air and Yellow Flags: Challenges in Predictive Analytics for Professional Racing**. Big Data, vol 2 issue 2, pages 97-112, June 20, 2014.
140. Şeyda Ertekin, Cynthia Rudin and Haym Hirsh. **Approximating the Crowd**. Data Mining and Knowledge Discovery, volume 28 issue 5-6, pages 1189-1221, September 2014.
    - Shorter versions appeared at NIPS Workshop on Computational Social Science and the Wisdom of Crowds, 2011, Proceedings of Collective Intelligence (CI) 2012, and Proceedings of the 2012 AAAI Fall Symposium on Machine Aggregation of Human Judgment, MAGG-2012.
141. Been Kim, Cynthia Rudin, and Julie Shah. **The Bayesian Case Model: A Generative Approach for Case-Based Reasoning and Prototype Classification**. Neural Information Processing Systems (NIPS), 2014.
142. Theja Tulabandhula, Cynthia Rudin. **On Combining Machine Learning with Decision Making**. Machine Learning (ECML-PKDD journal track), volume 93, Pages 33-64, 2014

143. Siong Thye Goh and Cynthia Rudin. **Box Drawings for Learning with Imbalanced Data**. Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2014.

144. Theja Tulabandhula and Cynthia Rudin. **Generalization Bounds for Learning with Linear, Polygonal, Quadratic and Conic Side Knowledge**. Machine Learning (ECML-PKDD journal track), December, 2014, pages 1-34.

    - Shorter version: Theja Tulabandhula, Cynthia Rudin. **Generalization Bounds for Learning with Linear and Quadratic Side Knowledge**. Proceedings of ISAIM 2014.

145. Jonathan Huggins and Cynthia Rudin. **A Statistical Learning Theory Framework for Supervised Pattern Discovery** Proceedings of SIAM Conference on Data Mining (SDM) 2014.

146. Been Kim and Cynthia Rudin. **Learning About Meetings**, Data Mining and Knowledge Discovery, (ECML-PKDD Journal track), volume 28 issue 5-6, pages 1134-1157, September 2014.

2013

147. Benjamin Letham, Cynthia Rudin and Katherine Heller. **Growing a List**. Data Mining and Knowledge Discovery (DAMI), ECML-PKDD journal track. volume 27, pages 372-395, 2013.

148. Tong Wang, Cynthia Rudin, Daniel Wagner, Richard Sevieri. **Learning to Detect Patterns of Crime**, Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 2013.

    - Ideas from this work were implemented by the NYPD by Alex Chohlas-Wood and E.S. Levine in their algorithm Patternizr, which operates live in New York City.

149. Cynthia Rudin, Benjamin Letham, and David Madigan. **Learning Theory Analysis for Association Rules and Sequential Event Prediction**. Journal of Machine Learning Research (JMLR), volume 14, pages 3385-3436, 2013.

    - Shorter version: Cynthia Rudin, Ben Letham, Ansaf Salleb-Aouissi, Eugene Kogan, and David Madigan. **Sequential Event Prediction with Association Rules**, Proceedings of the 24th Annual Conference on Learning Theory (COLT), 2011.

150. Benjamin Letham, Cynthia Rudin and David Madigan. **Sequential Event Prediction**. Machine Learning, volume 93, pages 357-380, 2013

151. Theja Tulabandhula and Cynthia Rudin. *Finalist, Data Mining Best Student Paper Competition, INFORMS 2012.* **Machine Learning with Operational Costs**. Journal of Machine Learning Research (JMLR), volume 14, pages 1989-2028, July 2013. Preliminary work is in the following conference paper.

    - Theja Tulabandhula and Cynthia Rudin. **The Influence of Operational Costs on Estimation**, Proceedings of the International Symposium on Artificial Intelligence and Mathematics (ISAIM), 2012.

2012

152. Tyler McCormick, Cynthia Rudin, and David Madigan. **Hierarchical Models for Association Rule Mining: A New Approach for Adverse Event Prediction in Clinical Trials**, Annals of Applied Statistics, volume 6, No. 2, pages 652–668, 2012.

153. Cynthia Rudin, David Waltz, Roger N. Anderson, Albert Boulanger, Ansaf Salleb-Aouissi, Maggie Chow, Haimonti Dutta, Philip Gross, Bert Huang, Steve Ierome, Delfine Isaac, Arthur Kressner, Rebecca J. Passonneau, Axinia Radeva, Leon Wu. **Machine Learning for the New York City Power Grid**, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 34, No 2, February 2012. *Spotlight Paper for the February 2012 Issue.*

154. Allison Chang, Cynthia Rudin, Mike Cavaretta, Robert Thomas and Gloria Chou. **Reverse-Engineering Quality Ratings**, Machine Learning: volume 88, issue 3, pages 369-398, 2012.

2011

155. Cynthia Rudin, Rebecca J. Passonneau, Axinia Radeva, Steve Ierome, and Delfina Isaac. **21st-Century Data Miners Meet 19th-Century Electrical Cables**, IEEE Computer, volume 44 no. 6, pages 103-105, June 2011.
(One of three articles featured on the cover of the magazine.)

156. Şeyda Ertekin and Cynthia Rudin. **On Equivalence Relationships Between Classification and Ranking Algorithms**, Journal of Machine Learning Research, volume 12, pages 2905–2929, 2011.

2010

157. Cynthia Rudin, Rebecca J. Passonneau, Axinia Radeva, Haimonti Dutta, Steve Ierome, and Delfina Isaac. **A Process for Predicting Manhole Events in Manhattan**. Machine Learning, volume 80, pages 1–31, 2010.

- Also oral presentation at ICML 2012

The following conference papers are also related to my projects on grid reliability.

- Rebecca J. Passonneau, Cynthia Rudin, Axinia Radeva, Ashish Tomar, Boyi Xie. **Treatment Effect of Repairs to an Electrical Grid: Leveraging a Machine Learned Model of Structure Vulnerability**, Proceedings of the KDD Applications in Sustainability (SustKDD) Workshop on Data Mining, 17th Annual ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2011.
- Dingquan Wang, Rebecca Passonneau, Michael Collins and Cynthia Rudin. **Modeling Weather Impact on a Secondary Electrical Grid**, 4th International Conference on Sustainable Energy Information Technology (SEIT-2014), 2014.
- Leon Wu, Timothy Teräväinen, Gail Kaiser, Roger Anderson, Albert Boulanger, and Cynthia Rudin. **Estimation of System Reliability Using a Semiparametric Model**, Proceedings of IEEE EnergyTech, 2011.
- Leon Wu, Gail Kaiser, Cynthia Rudin, and Roger Anderson. **Data Quality Assurance and Performance Measurement of Data Mining for Preventive Maintenance of Power Grid**, Proceedings of the KDD Workshop on Data Mining for Service and Maintenance (KDD4Service), 17th Annual ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2011.
- Leon Wu, Gail Kaiser, Cynthia Rudin, David Waltz, Roger Anderson, Albert Boulanger, Ansaf Salleb-Aouissi, Haimonti Dutta, and Manoj Poolery. **Evaluating Machine Learning for Improving Power Grid Reliability**, Proceedings of the ICML 2011 workshop on Machine Learning for Global Challenges, International Conference on Machine Learning, 2011.
- Axinia Radeva, Cynthia Rudin, Rebecca Passonneau and Delfina Isaac. **Report Cards for Manholes**, Proceedings of the International Conference on Machine Learning and Applications (ICMLA), 2009. *Best Poster Award*.
- Rebecca Passonneau, Cynthia Rudin, Axinia Radeva and Zhi An Liu. **Reducing Noise in Labels and Features for a Real World Dataset: Application of NLP Corpus Annotation Methods**, Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), 2009.
- Haimonti Dutta, Cynthia Rudin, Becky Passonneau, Fred Seibel, Nandini Bhardwaj, Axinia Radeva, Zhi An Liu, Steve Ierome, Delfina Isaac. **Visualization of Manhole and Precursor-Type Events for the Manhattan Electrical Distribution System**, Workshop on GeoVisualization of Dynamics, Movement and Change, 11th AGILE International Conference on Geographic Information Science, 2008.
- Boyi Xie, Rebecca J. Passonneau, Haimonti Dutta, Jing-Yeu Miaw, Axinia Radeva, Ashish Tomar, Cynthia Rudin. **Progressive Clustering with Learned Seeds: An Event Categorization System for Power Grid.** 24th International Conference on Software Engineering and Knowledge Engineering (SEKE 2012). Redwood City, CA. July 1-3, 2012.

2009

158. Cynthia Rudin. **The P-Norm Push: A Simple Convex Ranking Algorithm that Concentrates at the Top of the List**, Journal of Machine Learning Research, volume 10, pages 2233–2271, 2009.

- Shorter version: Cynthia Rudin. **Ranking with a P-Norm Push**. Proceedings of the Nineteenth Annual Conference on Learning Theory (COLT), pages 589 - 604, 2006.

An application of the P-Norm Push is described in this conference paper:

- Heng Ji, Cynthia Rudin, and Ralph Grishman. **Re-ranking Algorithms for Name Tagging**. In Proc. Human Language Technology conference - North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL) Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing, 2006.

159. Cynthia Rudin and Robert E. Schapire. **Margin-Based Ranking and an Equivalence Between AdaBoost and RankBoost**. Journal of Machine Learning Research, volume 10, pages 2193–2232, 2009.

    - Preliminary version: Cynthia Rudin, Corinna Cortes, Mehryar Mohri, and Robert E. Schapire. **Margin Based Ranking Meets Boosting in the Middle**. Proceedings of the Eighteenth Annual Conference on Learning Theory (COLT), pages 63 - 78, 2005.

2008 and before

160. Cynthia Rudin, Robert E. Schapire and Ingrid Daubechies. **Analysis of Boosting Algorithms Using the Smooth Margin Function**. Annals of Statistics, volume 35, number 6, pages 2723-2768, 2007.

    Preliminary material:

    - Cynthia Rudin, Robert E. Schapire, and Ingrid Daubechies. (2007) **Precise Statements of Convergence for AdaBoost and arc-gv**. In Proc. AMS-IMS-SIAM Joint Summer Research Conference: Machine Learning, Statistics, and Discovery, pages 131-145, 2007.
    - Cynthia Rudin, Robert E. Schapire, and Ingrid Daubechies. **Boosting Based on a Smooth Margin**. Proceedings of the Seventeenth Annual Conference on Computational Learning Theory, (COLT), pages 502-517, 2004.
    - Cynthia Rudin, Ingrid Daubechies, and Robert E. Schapire. **On the Dynamics of Boosting**. Advances in Neural Information Processing Systems (NIPS) 16, 2003.

161. Cynthia Rudin, Ingrid Daubechies, and Robert E. Schapire. **The Dynamics of AdaBoost: Cyclic Behavior and Convergence of Margins**. Journal of Machine Learning Research, 5 (Dec): 1557–1595, 2004.

    Preliminary material for this work appears partly within the NIPS paper below, and the open problem in COLT is from the JMLR paper:

    - Cynthia Rudin, Ingrid Daubechies, and Robert E. Schapire. **On the Dynamics of Boosting**. Advances in Neural Information Processing Systems (NIPS) 16, 2003.
    - Cynthia Rudin, Robert E. Schapire and Ingrid Daubechies. **Does AdaBoost Always Cycle?** JMLR: Workshop and Conference Proceedings, Published as a COLT Open problem, 2012.

162. Ryan Roth, Owen Rambow, Nizar Habash, Mona Diab, and Cynthia Rudin. **Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking**, The 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL/HLT), 2008.
163. Cynthia Rudin and Brian Spencer. **Equilibrium Ridge Arrays in Strained Solid Films**. Journal of Applied Physics, vol 86, pp 5530-5536, 1999.

Non-Peer-Reviewed Publications

164. Ezra Miller, Cynthia Rudin and Ingrid Daubechies. **Response to Letter to AMS Notices: Boycott collaboration with police**, September issue, AMS Notices, 2020. *We urge the mathematical community not to boycott interactions with police, but instead to work together to improve society.*

165. Matthew J Salganik, Lauren Maffeo and Cynthia Rudin. **Prediction, Machine Learning, and Individual Lives: an Interview with Matthew Salganik**. Harvard Data Science Review, 2020
166. Sarah Desmarais, Brandon Garrett, and Cynthia Rudin. **Risk Assessment Tools Are Not A Failed 'Minority Report'**. Perspectives, Law360, July 19, 2019, 5:50 PM EDT
167. Cynthia Rudin and David Carlson. **The Secrets of Machine Learning: Ten Things You Wish You Had Known Earlier to be More Effective at Data Analysis**. INFORMS TutORial, 2019.
168. Cynthia Rudin (with credit to Robin Smith). **Algorithms and Justice: Scrapping the Black Box.** The Crime Report, 2018.
169. Cynthia Rudin, David Dunson, Rafael Irizarry, Hongkai Ji, Eric Laber, Jeffrey Leek, Tyler McCormick, Sherri Rose, Chad Schafer, Mark van der Laan, Larry Wasserman, Lingzhou Xue. **Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society**. American Statistical Association whitepaper, `http://www.amstat.org/policy/pdfs/BigDataStatisticsJune2014.pdf`, 2014.
170. Cynthia Rudin and Kiri Wagstaff. **Machine Learning for Science and Society**, Machine Learning, (Introduction to the Special Issue on Machine Learning for Science and Society), volume 95, issue 1, April 2014, pp 1-9.
171. Cynthia Rudin. **Teaching "Prediction: Machine Learning and Statistics"**, Proceedings of the ICML Workshop on Teaching ML, 2012.
172. Peter Qian, Yilu Zhou, and Cynthia Rudin, **Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI)**, editors, 2011.
173. Cynthia Rudin and Miroslav Dudík, **Lecture Notes for the AMS Short Course on Statistical Learning**, editors, includes contributions by Robert E Schapire, Lawrence Saul, Lisa Hellerstein, Adam Tauman-Kalai, and John Lafferty, 2007.

Book

Intuition for the Algorithms of Machine Learning, Multimedia textbook, including videos, in progress, 2022. Free and publicly available. Introduces interpretable ML as part of introductory graduate ML.
`https://users.cs.duke.edu/∼cynthia/teaching.html`

Grants

Comparative effectiveness of EEG guided anti-seizure treatment in acute brain injury, NIH, PI Sahar Zafar, MGH, 12/1/2023-11/30/2028.

DMREF/Collaborative Research: Accelerated Discovery of Sustainable Bioplastics: Automated, Tunable, Integrated Design, Processing and Modeling, NSF, co-PI. 10/01/2023-09/30/2027. Amount of Award: $790,000 (with Cate Brinson, Eleftheria Roumeli, Linda Schadler, and Kayla Sprenger).

Novel Algorithm and Data Strategies to Detect and Predict Atrial Fibrillation for Poststroke Patients (NADSP), NIH (DHHS). co-PI, with Pi Xiao Hu (Emory). 3/10/2023-2/29/2024. Amount of Award: $551,830

Exploring the Whole Set of Sparse Explanations, DOE. PI. 09/01/2022-08/31/2024. Amount of Award: $400,000

FW-HTF-R: Interpretable Machine Learning for Human-Machine Collaboration in High Stakes Decisions in Mammography, NSF. PI, with co-PI Joseph Lo. 09/01/2022-08/31/2026. Amount of Award: $1,800,000

Patient-Focused Collaborative Hospital Repository Uniting Standards (CHoRUS) for Equitable AI, National Institutes of Health, co-PI, with PI Eric Rosenthal and several other co-PIs. 09/01/2022 - 08/31/2026. Amount of Award: $5,880,300

FAI: An Interpretable AI Framework for Care of Critically Ill Patients Involving Matching and Decision Trees, NSF & Amazon. PI, with co-PIs Alex Volfovsky and Sudeepa Roy. 07/01/2022-06/30/2025. Amount of Award: $1,000,000

Center for Virtual Imaging Trials. NIH/NIBIB. Co-lead for Project 3, with PIs Ehsan Samei and Joseph Lo. 4/01/21-12/31/25. Technology Research and Development (TR&D) Project 3. Amount of Award: $6,371,435 (total) $2,620,393

(Project 3).

MedX: Interpretable Deep Learning Models for Better Clinician-AI Communication in Clinical Mammography, 11/1/2021-10/31/2022. Internal Duke Award, with Joseph Lo, Amount of Award: $50,000

A Machine Learning Framework for Understanding Impacts on the HIV Latent Reservoir Size, Including Drugs of Abuse, National Institute on Drug Abuse (NIDA), PI, with co-I's David Murdoch, Nilu Goonetilleke and Nancie Archin. 9/30/2021-9/30/2026. Amount of Award: $2,284,375

Research on Explainable AI on Multiple Models, Fujitsu, PI, 07/16/2021-7/16/2024. Amount of Award: $272,307

EAGER: Creating an Unsupervised Interpretable Representation of the World Through Concept Disentanglement, NSF, PI, 11/01/2021-10/31/2024. Amount of Award: $169,345

NSF Workshop on Seamless/Seamful Human-Technology Interaction, NSF, PI, 09/01/2021-08/31/2022. Amount of Award: $49,777

FAIR Data and Interpretable AI Framework for Architectured Metamaterials, DOE, co-PI, with PI Cate Brinson, and co-PI Chiara Daraio, 9/1/2020-8/31/2023. Amount of Award: $870,580

NRT-HDR Harnessing AI for Understanding & Designing Materials (aiM), NSF, co-PI, with PI Cate Brinson, co-PIs David Banks, Stafano Curtarolo, Johann Guilleminot, 09/01/2020-08/31/2025. Amount of Award: $2,252,971

NSF-HDR TRIPODS: Innovations in Data Science, CCF-1934964 co-PI, with PI Sayan Mukherjee, co-PI's Robert Calderbank, Rong Ge, and Jianfeng Liu, 10/1/2019-9/30/23. Amount of Award: $1.5 million

Utilizing Key Past Experiences from Large Datasets to Make Better Prediction in Multi-Class Settings, co-PI with Ramin Moghaddass, Amazon AWS Machine Learning Research Awards program, 2018. Amount of Award: $20,000

An Integrated Nonparametric Bayesian and Deep Neural Network Framework for Biologically-Inspired Lifelong Learning (co-PI, with PI Katherine Heller, and other co-PI's Lawrence Carin, David Dunson, Nicolas Brunel, Tamara Broderick, Joshua Tenenbaum, Michael Jordan, Thomas Griffiths), DARPA, 2017-2020. Amount of Award: $4,388,119

Collaborative Research Framework: Data: HDR: Nanocomposites to Metamaterials: A Knowledge Graph Framework, OAC-1835782 (co-PI, with PI Cate Brinson, and other co-PI's Chiara Daraio, Linda Schadler, Deborah McGuinness, and Wei Chen), NSF, 11/1/2018-10/31/2023. Amount of Award: $2,590,810

Lord Foundation "Duke's Super Superresolution Team (Continued)" (Single PI) 2019-2020. Amount of Award: $24,000

Lord Foundation "Duke's Super Superresolution Team" (Single PI) 2018-2019. Amount of Award: $32,000

Duke Energy Initiative "Enabling Better Energy Decisions Through Better Interpretable Causal Inference Methods for Personalized Treatment Effects" Duke University Energy Initiative Energy Research Seed Fund (ERSF). (PI, with co-PI's Alex Volfovsky and Sudeepa Roy), 2018-2021. Amount of Award: $40,500 Stage I, $40,500 Stage II

QuBBD: Collaborative Research: Matching Methods for Causal Inference: Big Data and Networks, 1R01EB025021-01, sponsored by DHHS, PHS, NIH, NIBI&B" (co-Investigator, with PI Alexander Volfovsky, co-Investigators Sudeepa Roy and Allison Aiello). 2017-2020. Amount of Award: 849K$($514K to Duke)

Alfred P. Sloan Foundation "Interdisciplinary Energy Data Analytics Ph.D. Fellows Program," (co-PI, with Brian Murray, William Pizer, and Kyle Bradbury). 2018-2020. Amount of Award: $225,000

Laura and John Arnold Foundation "Interpretable Machine Learning for Pre-Trial Risk Analysis," 2017-2020. (Single PI) Amount of Award: $87,507

Duke Institute for Health Innovation "Pallialytics: Using analytics to inform a palliative care population health management intervention" (co-PI with Jonathan Fischer, Leslie Alabi, Eugenie Komives from Duke Medical). 2017-2018. Amount of award: $55,500

2016 Adobe Digital Marketing Research Award, "Compute-intensive causal machine learning models for finding customer segments" (Single PI). 2016. Amount of Award: $50K

MIT-Lincoln Labs "Adaptable, Interpretable, Machine Learning" (co-PI with Jonathan Su from MIT-LL), 2016-2019. Amount of Subaward: $542K

Xerox Research "Causal Inference-related Research Threads in the Prediction Analysis Lab, continued" (co-PI with Theja Tulabandhula), 2016. Amount of Award: $30K

DARPA "Foundations of Sequential Learning" (co-PI with Ron Parr and Kamesh Munagala), April 2016 - Jan 2017. Amount of Award: $242K

Philips "Algorithms for Interpretable Risk-Scoring" (Single PI), July 2017-December 2017. Amount of Award: $122K

Xerox Research "Causal Inference-related Research Threads in the Prediction Analysis Lab" (co-PI with Theja Tulabandhula), 2015. Amount of Award: $30K

Philips "Self-Learning Systems and User-Behavior Modeling" (Single PI), June 2015-May 2016. Amount of Award: $240K

Army Research Office "Uncertainty Quantification for Unobserved Variables in Dynamical Systems and Optimal Experimental Design" (Single PI), Spring 2015. Amount of Award: $50K

Big Data Seed Grant, MIT Big Data Initiative "Interpretable, Scalable and Causal Models from Machine Learning" (Single PI), Spring 2015. Amount of award: 1 semester RA

Accenture and MIT Alliance in Business Analytics "Creating a Model of the Usual State of a Machine" (Single PI), March 20 2015 - March 20 2016. Amount of award: $100,000

Wistron Corporation "Interpretable Predictive Models from Machine Learning" (Single PI), September 4, 2014 - August 31, 2016.
Amount of award $ 300,000

Accenture and MIT Alliance in Business Analytics "Big Data Analysis for Plant and Commercial Optimization," June 15 2013 - June 14 2014. Amount of award: $ 100,000

Ford Racing "Predictive Analytics for Racing" (Single PI),
September 1 2012 - August 31 2013. Amount of award: $75,000
September 1 2013 - December 31 2014. Amount of award: $75,000
December 31 2014 - December 31 2015. Amount of award: $157,000

Siemens Energy, CKI University Innovation Initiative. "CKI Proposal: Augmented Data-Driven Diagnosis using Physical Models" (Single MIT PI with co-PIs at Siemens), June 1, 2013 - May 31, 2016
Amount of award: $350,000

Siemens Energy, CKI University Innovation Initiative. "CKI Proposal: Incorporating Prediction Analysis into XHQ" (Single MIT PI with co-PIs at Siemens), June 1 2013 - May 31 2015.
Amount of award: 150,000 euro ≈ $200,000

Ford - MIT Alliance. "Develop advanced in-vehicle SYNC advertisement features to target drivers based on their context information, system interaction and past choices" (Single PI), July 1 2012 - August 31 2014.
Amount of award: $298,000

MIT Sloan Research Fund. "Predictive Models for Highly Imbalanced data," (Single PI), Amount of award: $20,000

Solomon Buchsbaum Research Fund. "A New Foundation for Statistical Decision-Making," June 2 2011-present. (Single PI), Amount of award: $50,000

NSF-CAREER IIS-1053407. "New Approaches for Ranking in Machine Learning," September 1 2011 – August 31 2018. Amount of award: $480,000

MIT Lincoln Labs. "A Knowledge Discovery Framework for Threat Identification (Phase III)" (co-PI with Shirish Ranjit from MIT-LL) 7/1/2013-6/30/2014.
Amount of award: $100,000

MIT Lincoln Labs. "A Knowledge Discovery Framework for Threat Identification (Phase II)" (co-PI with Shirish Ranjit from MIT-LL) 7/1/2012-6/30/2013.
Amount of award: $100,000

MIT Lincoln Labs. "A Knowledge Discovery Framework for Threat Identification" (co-PI with Shirish Ranjit from MIT-LL and Regina Barzilay from MIT), 6/1/2011- 5/31/2012.
Amount of award: $83,000

MIT Intelligence Initiative. "Combining human and machine predictions using "boosting" algorithms" (co-PI with Thomas Malone and Patrick Winston) November 1 2010 - October 1 2011.
Amount of award: 0.5 postdoctoral research fellowship

MIT Energy Initiative (MITEI) Seed Fund Program. "A Novel Framework for Electrical Grid Maintenance" (Single PI) May 17 2010 - May 16 2012.
Amount of award: $150,000

Ford - MIT Alliance. "Achieving Top Quality Ratings with Minimal Cost" (Single PI) July 1 2010 - June 20 2011.
Amount of award: $106,013

Con Edison Company of New York. "Manhole Events and Secondary System - Machine Learning Project"

- Secondary System Project, Manhattan Backbone, December 1, 2011-December 31, 2013, Amount of subaward to MIT (Single PI): $300,129
- Secondary System Project, Manhattan Corollary, May 1, 2011 - May 31, 2012, Amount of subaward to MIT (Single PI): $93,056
- Update Manhattan Consolidated Database and Ranking Model through 2009, and Analyze 2004-2005 Secondary Inspections Data, July 1 2010 - December 31 2010, Amount of subaward to MIT (Single PI): $81,661
- Phase 2, Application to B, Q, and X, December 1 2009 - May 31 2010, Amount of subaward to MIT (Single PI): $81,899
- Phase 1, Application to B, Q, and X, January 2009 - June 2009, Co-PI , Amount: $464,127
- Phase 4, July, 2008 - December, 2008, Co-PI , Amount: $413,947
- Phase 3, January, 2008 - June, 2008, Co-PI , Amount: $347,104
- Phase 2, August, 2007 - December, 2007, Co-PI, Amount: $486,296
- Phase 1, March, 2007 - July, 2007, Co-PI , Amount: $339,251

National Science Foundation. Postdoctoral Research Fellowship in Biological Informatics, Grant DBI-0434636, March 2005 - February 2007.
Amount of award: $120,000

Honors (not including awards listed with papers above)

Gilbert, Louis, and Edward Lehrman Distinguished Professorship in Computer Science, Duke University, 2024-present
INFORMS Senior Member, 2024-present

Bell Labs Prize, Second Place, 2023. Recognizes game-changing innovations in science, technology, engineering and mathematics. 107 teams competing, 10 semifinalists, 5 finalists.

David and Janet Vaughan Brooks Teaching Award, Duke University, 2023. Awarded by the Trinity College Arts and Sciences Council based on equitable and inclusive teaching, sparking excitement about learning, encouraging a deep dive into disciplinary ways of thinking, making connections beyond the courses they teach, and teaching innovations. Highest teaching award at Duke, 4 teaching awards per year across all departments.

Grand Prize Winner, A-μ-Sing Competition, Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), 2023. This award was based on three bluegrass songs that I wrote about machine learning for my class.

Named by AIo as one of the top 10 Women in the World of AI in 2023

7th International Workshop on Health Intelligence (W3PHIAI-23, AAAI-23 workshop), Winner of Hackathon ($1K awarded to student Srikar Katta), 2023. Objective: This hackathon challenged the AI community to design an optimal age predictor. There was no restriction on the type of data used for this challenge, and we used sleep EEG to predict biological ”brain” age.

Bass Chair (Earl D. McLean, Jr. Professor), Duke University, 2022-2024

Chron15 Pioneer 2021-2022, Duke Chronicle, 2022

Bass Society of Fellows, Duke University, Inducted 2022

Terng Lecturer, Women in Mathematics Program, Institute for Advanced Study, Princeton, 2022

2022 Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity from the Association for the Advancement of Artificial Intelligence (AAAI). This is the most prestigious award in artificial intelligence. This award, similar only to world-renowned recognitions, such as the Nobel Prize and the Turing Award, carries a monetary reward at the million-dollar level.

Guggenheim Fellow, 2022

Fellow of the Association for the Advancement of Artificial Intelligence (AAAI), 2022-present

Thomas Langford Lecture Award, Duke University, 2019-2020

Fellow of the American Statistical Association, 2019-present

Fellow of the Institute of Mathematical Statistics, 2019-present

Duke AI for Art competition (with students Alina Barnett, James Hoctor, Chaofan Chen, and Oscar Li), second place, 2019

Faculty associate, Berkman Klein Center for Internet and Society at Harvard University, 2015-2019.

2016 Adobe Digital Marketing Research Award, 2016

Named by Businessinsider.com as one of the 12 most impressive professors at MIT in 2015

National Science Foundation CAREER Award, 2011

“Top 40 Under 40” business school professors of *Poets & Quants*, 2015. (Published in Forbes magazine.)

Nominated for Outstanding UROP Mentor Award, UROP (undergraduate research opportunities) Program, MIT, 2012

Nominated for 2012 Sloan Excellence in Teaching Awards, MIT, 2012

Second Place for Phase 1 in the ICML Exploration and Exploitation 3 Challenge, 2012. Goal is to design a recommender system with high click through rates for Yahoo! Front Page News Article Recommendations
teammates: Virot Ta Chiraphadhanakul and Edward Su

National Science Foundation Postdoctoral Research Fellowship in Biological Informatics, 2005-2007

University at Buffalo College of Arts and Sciences Outstanding Senior Award in Sciences and Mathematics 1999, one per year at the university (also Department of Physics Outstanding Senior 1998, Department of Mathematics Outstanding Senior 1999, Department of Music Outstanding Senior 1999)

Barry J. Goldwater Scholarship, 1997-1998

State University of New York Chancellors Award for Student Excellence 1997, 1999

Dr. Stanley T. Sekula Memorial Scholarship, University at Buffalo Physics Department, 1996, 1997

Hildegarde F. Shinners Prize, 1999, Phi Beta Kappa award to recognize mathematics thesis and music thesis

Phi Beta Kappa, inducted 1997

UB Music Department Scholarships, 1994, 1995, 1996

Intellectual Breath and Liberal Knowledge Award, UB Honors Program, 1999

Second Place-National Winner of the 1993 Young Inventors’ and Creators’ Competition in the Copyright Category of Popular Music Composition, sponsored by the Foundation for a Creative America. (This came with a congratulatory letter signed by Vice President Al Gore!)

Testimony

- Deposition in Case: Flores v. Stanford 7:18-cv-02468 — U.S. District Court for the Southern District of New York, December 14, 2022. Expert report, May 30, 2022

Media (Selected)

Squirrel AI Award:

"Duke Computer Scientist Wins $1 Million Artificial Intelligence Prize, A 'New Nobel' " AAAI press release, October 12, 2021

"Duke Professor Recognized for Clarifying AI Decision Making" by John McCormick, Wall Street Journal, Lead article in AI section, October 14, 2021

"She won a $1 million prize for predicting which manholes would explode" by James Barron, New York Times, New York Today section, October 12, 2021 (This title is misleading.)

AI Regulation:

"Biden's Executive Order on AI Is a Good Start, Experts Say, but Not Enough" by Lauren Leffer, Scientific American, October 31, 2023

"What You Need to Know About Biden's Sweeping AI Order: The executive order covers AI safety, algorithmic bias, and privacy" by Eliza Strickland, IEEE Spectrum, October 10, 2023

NYT quote in "How Easy Is It to Fool A.I.-Detection Tools?" by Tiffany Hsu, New York Times, June 28, 2023

CNBC International (newsroom in Singapore), interview on live television, May 29 2023

"Senators reaffirm worries about AI during hearing with ChatGPT founder," Ahtra Elnashar, The National Desk, WJLA (ABC News 7 - based in Virginia), May 16th, 4:00pm EST, 2023

"Duke professor - a critic of 'runaway train' AI development - is named one of the field's top women", WRAL TechWire, April 6, 2023

WRAL TechWire interview: "Artificial intelligence industry is out of control, requires regulation, Duke researcher warns," March 6, 2023

CNN quote in "Experts are warning A.I. could lead to human extinction. Are we taking it seriously enough?" May 30, 2023

Machine Learning Interpretability Interviews:

LinkedIn Live Conversation with Scott Zoldi from FICO, March 8, 2023

"What Do Conspiracy Theories And AI Explainability Have In Common?" by Kareem Saleh, Forbes, May 4, 2021

"Meet Cynthia Rudin–A Champion of Interpretable Machine Learning," by Sam Behseta & Michelle Dunn, Chance Magazine, American Statistical Association, April 22, 2020

"Rise of Robot Radiologists," by Sara Reardon, Nature, Innovations In, December 18, 2019

BBC radio, Digital Planet, December 10, 2019 (discussing "This Looks Like That")

"Machine vision that sees things more the way we do is easier for us to understand" MIT Technology review, Artificial Intelligence, Dec 6 2019

Computer Vision:

"Accurate Neural Network Computer Vision Without the 'Black Box': Duke team disentangles neural networks to understand how they see the world," December 15. 2020

"Artificial Intelligence Makes Blurry Faces Look More Than 60 Times Sharper: This AI turns even the blurriest photo into realistic computer-generated faces in HD": by Robin A. Smith, Duke Today, June 12, 2020 - published on DukeToday, covered by Newsweek.com, Independent.co.uk and other media outlets.

"Making Blurry Faces Photorealistic Goes Only So Far," by Mark Anderson. IEEE Spectrum, June 23 2020

Crime Data Mining:

"Possible to predict recidivism? Here's how...," The Docket, MSNBC (Live TV), Tuesday May 19, 2015.

"Crime-Fighting Computer Code from Cambridge Police and MIT," *WBUR Radio Boston* (National Public Radio), Tuesday August 13 2013

"Cambridge police look at math to solve crimes," *Boston Globe*, Metro Section, front page, Sunday August 4 2013

"Statistician enlisted to fight crime by numbers," *The Times of London*, US & Americas section, Tuesday August 6 2013

"Predictive Policing: Using Machine Learning to Detect Patterns of Crime," *WIRED Innovation Insights*, August 22, 2013

Meetings Analysis:

"At Work: Just Say 'Yeah'," *Wall Street Journal*, Business section, on page B8 in the U.S. edition, June 19, 2013

"How to be effective at meetings? Say 'yeah'," *Toronto Star*, Business section, June 28, 2013

"Researchers discover the key to persuasion," *ABC News*, consumer report blog / business, June 24, 2013

Energy Grid:

My work discussed in book The Alignment Problem: Machine Learning and Human Values 1st Edition by Brian Christian, W. W. Norton and Company, 2020.

"Why Manhole Explosions Happen in the Summer," *NBC News*, Business/Energy section, August 19, 2015

"New York's Exploding Manhole Covers Pose Unexpected Winter Hazard," *Reuters*, appeared in *New York Times*, and *MSN.com*, February 28, 2015

My work discussed in book Big Data: A Revolution that Will Transform how we Live, Work, and Think, by Victor Mayer-Schönberger and Kenneth Cukier, Houghton Mifflin Harcourt Publishing Company, 2013

Analytics Magazine, INFORMS. Headlines: Innovative Applications in Analytics Award, April 18, 2013

"Machine Versus Manhole," *ScienceNews, U.S. News and World Report, WIRED Science, Slashdot, Discovery News / Discovery Channel*, and others, July 8-9 2010

Information Retrieval / Building New Search Engines:

Radio segment about my work on Growing a List. "A New Way to Google," *Boston Public Radio*, show on innovation at 12:45pm-1pm, hosted by Kara Miller, October 9, 2012

Health and Interpretable Predictive Models:

"New AI tool from Duke researchers reads EEGs, could improve treatment of seizures," by Freya Gulamali, The News & Observer, June 14, 2024

"How Can Doctors Be Sure A Self-Taught Computer Is Making The Right Diagnosis?," by Richard Harris, All Things Considered, NPR, April 1, 2019

"Do You Zone Out? Procrastinate? Might Be Adult ADHD," by Rebecca Hersher, NPR, April 5 2017

"Algorithms Learn From Us, and We've Been Bad Parents," by Bahar Gholipour, Mach Technology, NBC News, March 10 2017, 2:17 PM ET

"New Computer Tool Can Predict Dementia From Your Simple Drawings" Popular Science, August 13, 2015

"Digital Pen is Better Dementia-Prediction Tool than a Doctor" WIRED Magazine, August 17, 2015

"Computers that teach by example: New system enables pattern-recognition systems to convey what they learn to humans." MIT News (also front page of MIT main website), December 5-10, 2014

"New Statistical Model Lets Patient's Past Forecast Future Ailments," Science News section, *Science Daily*, June 9, 2012

Other topics:

Quoted in "AI Is Getting Powerful. But Can Researchers Make It Principled?" by Mordechai Rorvig, Scientific American, April 4, 2023.

Scientific Sonnets: Duke team wins competition for poetry-generating algorithm, Duke Chronicle, Dec 27, 2018

Article discussing the whitepaper effort I led: AmstatNews: The Membership Magazine of the American Statistical Association News. Cover Story, *Statistical Scientists Advance Federal Research Initiatives*, July 1, 2014.

Article about my work: "How to Improve Product Rankings," *Businessweek*, B-School Research Briefs section, March 9, 2012

Interview: "Should There Be Enforceable Ethics Regulations on Generative AI?" InformationWeek, article by Joao-Pierre S. Ruth, March 14, 2023

Op-Eds:

Cynthia Rudin. **Wreaking Havoc on Academic Publishing** (chatbots are awash in stolen academic data from reviewers). Inside Higher Ed, May 14, 2024.

Brandon Garrett and Cynthia Rudin. **Seeking To Regulate AI Is a Good Start. Next, Tackle the Secretive Way Government Uses It**. The Messenger, November 2, 2023.

Cynthia Rudin, Zhicheng Guo, Cheng Ding and Xiao Hu. **How good are AI health technologies? We have no idea**, STAT, First Opinion, Oct 11, 2023

Brandon Garrett and Cynthia Rudin. **What the Supreme Court Doesn't Understand About AI**, The Messenger, June 3, 2023

Cynthia Rudin and Lance Browne. **A Reasonable Right to Biometric Privacy.** (They retitled it to this title: "That selfie you posted on Instagram? Companies are using it in unethical ways.") Published in: Charlotte Observer, Raleigh News & Observer, Durham Herald-Sun, and Las Vegas Sun, March 27, 2023

Cynthia Rudin. **How dangerous is AI? Regulate it before it's too late**, The Hill, February 8, 2023

Cynthia Rudin. **'The Marriage Pact' and the risks we take with data**, The News&Observer, February 28, 2021

Cynthia Rudin. **No More Excuses. Make Data More Accessible.** Washington Post, Opinions, part of collection "We need smart solutions to mitigate the coronavirus's impact. Here are 46." June 18, 2020

Professional Societies and Government Committees

National Artificial Intelligence Advisory Committee Law Enforcement Subcommittee (NAIAC-LE), 2023-2026

Committee on Facial Recognition Technology: Current Capabilities, Future Prospects, and Governance, National Academies of Science Engineering and Medicine, 2022-2024

Member, Executive Committee, ACM SIGKDD, 2021-present

US Department of Labor Technical Working Group on Automated Approaches for Data Catalogs, May 17, 2023

Associate Director, Statistical and Applied Mathematical Sciences Institute (SAMSI), 2018-2021

Chair of Committee to Choose the First Editor-in-Chief of the new INFORMS Journal on Data Science, INFORMS, 2019-2020

Judge for Edelman Award, INFORMS 2019-2020, 2020-2021

Technology Strategy Committee, INFORMS, 2019-2022

Member of Committee for Computing Community Consortium (CCC) for Interaction for AI / Roadmap for Future AI, January 2019

Chair, Section on Statistical Learning and Data Mining, American Statistical Association, 2017-2018.

Member of Committee on Applied and Theoretical Statistics (CATS), National Academies of Sciences, Engineering, and Medicine, 2016-2022

Member of Committee on Law and Justice (CLAJ), National Academies of Sciences, Engineering, and Medicine, 2017-present

Councilor of the AAAI. 2017-2020

Chair, INFORMS Data Mining Section, 2015-2016 (Vice Chair for 2014-2015, Council Member 2017-2018, Council Member, 2011- 2013).

Member of Committee on Analytical Research Foundations for the Next-Generation Electric Grid, and author of consensus report: "Analytical Foundations for the Next Generation Electric Grid," National Academies of Sciences, Engineering and Medicine, 2014-2016.

Bureau of Justice Assistance Criminal Justice Technology Forecasting Group (BJA CJTFG), United States Department of Justice, 2014-2016.

DARPA Information Science and Technology (ISAT) study group (faculty advisory board of DARPA), 2014-2018.

American Statistical Association Committee on Funded Research, 2015-2018.

MIT Energy Initiative, Energy Education Task Force, 2013-2014.

Activities

Events Organized

Seamless/Seamful Human Computer Interaction, NSF workshop, co-organizer, May 17 and 20, 2021.

FAIF: Fair AI in Finance, NeurIPS workshop, co-organizer, 2020.

Self-Supervised Learning – Theory and Practice, NeurIPS workshop, co-organizer, 2020.

Triangle Machine Learning Day, lead organizer, co-organized with David Banks, Jeremy Freeman, and Ted Enamorado, September 20, 2019.

SAMSI Deep Learning Program, co-organizer and leader of working group, 2019

Conference co-Chair, Conference on Statistical Learning and Data Science / Nonparametric Statistics, co-Chair with Annie Qu, American Statistical Association, June 4-6, NYC, 2018

Triangle Machine Learning Day, lead organizer, co-organized with Jade Vinson, Richard Lucic, and Kirsten Shaw, April 3, 2018.

Member of planning committee, SAMSI Program on Statistical, Mathematical, and Computational Methods for Precision Medicine 2016-2018.

TAMALE: Toolkit of Algorithms for Machine Learning, DARPA ISAT workshop, co-organized with Margo Seltzer, March 2018.

Judge for INFORMS Data Mining Best Paper competitions, INFORMS, 2018.

Member of organizing committee, Conference on Statistical Learning and Data Science, UNC Chapel Hill. 2016.

What if: Machine Learning Models for Causal Inference, DARPA ISAT workshop, co-organized with Dustin Tingley, February 2016.

The Cassandra Problem: Building Trust in Predictive Models, DARPA ISAT workshop, co-organized with Carla Brodley and Stephen Boyd, April 2015.

Invited Session: The Fifth V in "Big Data" is *Variables*, co-organizer with Tyler McCormick, Joint Statistical Meetings, 2015.

Topic Contributed Session: Predictive Policing, organizer, Joint Statistical Meetings, 2014.

Judge for Statistical Learning and Data Mining Best Student Paper competition, American Statistical Association, 2015.

Judge for 2015 INFORMS Innovative Applications in Analytics Award, 2014-2015.

Workshop on Data Analytics: Challenges in Big Data for Data Mining, Machine Learning and Statistics organizer, MIT CSAIL Big Data, March 26, 2014.

The ISBIS (International Society of Business and Industrial Statistics) 2014 and SLDM (Statistical Learning and Data Mining section of the American Statistical Association) Meeting on Data Mining in Business and Industry, Program Committee, June 9-11, 2014.

Statistical Analysis and Data Mining (a journal of the American Statistical Association), committee to choose the next editor-in-chief, 2014-2015.

ECML/PKDD 2013 Workshop on "DARE: Data Analytics for Renewable Energy Integration", Technical program committee member, 2013.

Workshop on Hospital Readmission Prediction and Clinical Risk Management (HRPCRM) at IEEE International Conference on Healthcare Informatics (ICHI) 2013, program committee member, organizers are John Cromwell and Si-Chi Chin

Session on Smart Grid Data Analytics (SGDA) at International Conference on Smart Grid and Clean Energy Technologies (ICSGCE) 2013, co-chair with Zeyar Aung

Dagstuhl Seminar on Preference Learning, co-organizer, 5 day seminar, 45 participants, Germany, March 3th to March 7th, 2014.

IMS/ASA Spring Research Conference, organizer of Machine Learning session, Harvard University, June 14th, 2012.

New England Machine Learning Day, Co-organizer, Microsoft Research New England, May 16th, 2012.

New England Statistics Symposium (NESS), organizer of Machine Learning session, Boston University, April 21, 2012.

Collective Intelligence 2012 (CI 2012), Local Arrangements Chair, 2012

INFORMS Data Mining and Health Informatics Workshop (DH-MI), co-organizer, 2011

MIT Energy Initiative, Organizing Committee for the MITEI Seminar Series, member, Fall 2010- Spring 2012

AMS Short Course on Aspects of Statistical Learning, organizer, 2007 AMS joint math meetings, New Orleans, January 3-4, 2007

AMS/AWM/MAA Special Session on Mathematical Results and Challenges in Learning Theory, session organizer, AMS joint math meetings, San Antonio Texas, January 12-15, 2006

Program for Women in Mathematics, Institute for Advanced Study, Program Committee Member, 2003-2006. Women in Science Seminar organizer, 2004, 2005, 2006. TA for the Wavelets course in 2002. Discussion group organizer, 2001. Research seminar speaker 2003. Poster session organizer 2003, Panel discussion organizer 2003, Panelist 2007

PACM Conference Princeton University, organizer, 2002-2004, speaker 2005

Editorial Responsibilities

Associate Editor for Harvard Data Science Review, 2019 - present

Associate Editor for Management Science, in the Big Data Analytics department, 2018 - 2024

Associate Editor for INFORMS Journal on Data Science, 2020 - 2023

Associate Editor for Journal of Quantitative Criminology, 2021 - 2022

Action editor: 2013 - 2017, (Editorial board member: June 30 2010 - June 30 2013), Machine Learning Journal

Action Editor: Statistical Analysis and Data Mining (SAM, a journal of the American Statistical Association), 2012-2017

Editor for Special Issue on Sports Analytics for Statistical Analysis and Data Mining (SAM, a journal of the American Statistical Association), co-editor with Theja Tulabandhula, 2015

Editorial Board, Journal of Artificial Intelligence Research (JAIR), July 2014 - June 2017

Member of Guest Editorial Board for ECMLPKDD 2014 journal track

Editor for Special Issue on Machine Learning for Science and Society, co-editor with Kiri Wagstaff, 2012-2013

Editorial board member: Journal of Machine Learning Research, 2012 - present

Program Committee Memberships

NIPS Workshop on Machine Learning for Healthcare ML4HC (2016, 2017, 2018), ECML-PKDD Workshop on Social Good (2016), ICML Workshop on Human Interpretability for Machine Learning (2016), ECML-PKDD Workshop on Data Science for Social Good (2016), SDM senior pc (2016), ICDM area chair (2015), AAAI senior pc (2016), NIPS area chair (2015), Visual Data Science (2015), IJCAI senior pc (2015), ICML area chair (2015) ICDM area chair (2014), IEEE Big Data (2013), NIPS area chair (2013), AAAI (2013), ICML area chair (2013), NIPS area chair (2012), ECML-PKDD (2012), ICML (2012), COLT (2011), IJCAI (2011), ECML-PKDD (2010), ICML (2009), ICML (2006), AAAI (2006), KDD (2005)

Referee Assignments

Journal Reviewing:

Biometrics, Harvard Data Science Review, Journal of Artificial Intelligence Research (JAIR), Journal of Quantitative Criminology (JOQC), Data Mining and Knowledge Discovery (DAMI), Journal of Machine Learning Research (JMLR), Machine Learning Journal, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Mathematical Programming, Journal of Computational Statistics and Data Analysis, Communications on Pure and Applied Mathematics (CPAM), Management Science, International Journal of Renewable Energy Research (IJRER), Annals of Operations Research, Operations Research , Recent Patents on Computer Science, The Lancet, Scientific Reports - Nature, Science Advances, Decision Support Systems, AI Magazine

Conference and Other Reviewing: NeurIPS/NIPS (Neural Information Processing Systems) 2020, 2019, 2018, 2017, 2011, 2010, 2009, 2008, 2007. IEEE Computer 2019, ICDM (International Conference on Data Mining) 2011, Grant Proposal Review for Natural Sciences and Engineering Research Council of Canada 2016, Book Proposal for Springer 2013, COLT (Conference on Learning Theory) 2012, COLT 2010, COLT 2005, AISTATS (Conference on Artificial Intelligence and Statistics) 2012, AISTATS 2010, ICMLA (International Conference on Machine Learning and Applications) 2011, Book Proposal for Manning Publications 2010, ALT (Algorithmic Learning Theory) 2010, ICML (International Conference on Machine Learning) 2010, NIPS Ranking Workshop 2009, ACM SIGKDD (Conference on Knowledge Discovery and Data Mining) 2009, Applied and Computational Harmonic Analysis Journal, 2006, Conference on Machine and Statistical Learning: Prediction and Discovery 2006 , Reviewer for MRC Biostatistics Unit at the University of Cambridge 2023

Presentations

Keynotes/Invited/Plenary for Notable Conferences

Conference on Statistical Practice, American Statistical Association, Keynote, February 28, 2024

SPIE Medical Imaging, Opening Plenary, February 18, 2023

International Conference on Data Mining (ICDM), Keynote, November 30, 2022

INFORMS (Institute for Operations Research and the Management Sciences), Plenary, October 17, 2022

SDM (SIAM International Conference on Data Mining), Keynote, April 29, 2022

AAAI (Association for the Advancement of Artificial Intelligence), Plenary talk for Squirrel AI Award, February 24, 2022

Nobel Conference, Gustavus Adolphus College, October 4-6, 2021

CODE (Conference on Digital Experimentation), MIT, November 2, 2019

DSAA (IEEE International Conference on Data Science and Advanced Analytics), Washington DC, October 7th, 2019

KDD (25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining), Anchorage, Alaska, August 8th, 2019

Women in Data Science Cambridge Conference (organized by Harvard, MIT, Microsoft Research New England, Stanford) March 4, 2019

ECML-PKDD (The European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases), Dublin, September 11, 2018

Females Excelling More in Math, Engineering, and Science (FEMMES), keynote for capstone event, 200 girls grades 4-6, Duke University, February 17, 2018

ISAIM (International Symposium of Artificial Intelligence and Mathematics), January 3, 2018

MLHC (Machine Learning for Healthcare), August 18, 2017

AISTATS (Artificial Intelligence and Statistics), April 22, 2017

The Frontiers of Machine Learning, Forum on Machine Learning, National Academy of Sciences and The Royal Society, Jan 31-Feb 1, 2017

FAT-ML (Fairness, Accountability and Transparency in Machine Learning), November 18, 2016

KDD (20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining), August 24-27, 2014

Other Conference Invited Presentations

(This list does not include regular or invited talks at annual meetings such as JSM, INFORMS, SLDM/SLDS, American Society for Criminology, Atlantic Causal Inference Conference, which I participate in most years.)

Responsible AI for Health Symposium, Johns Hopkins University, August 29, 2024

Deep Learning School @UCA - France, Université Cote d'Azur, July 1, 2024

Duke AI Day, opening keynote, June 7, 2024

Explainable AI for Speech and Audio (XAI-SA) ICASSP Workshop, Plenary, April 15, 2024

Winter School for Ethics of Neuroscience and AI, invited talk, March 1, 2024

XAI4Sci, Explainable AI for Sciences Workshop, AAAI, invited speaker, February 26, 2024

CanQueue (Conference on Queueing Theory), keynote, August 25, 2023

Disney Data & Analytics Conference (DDAC), featured speaker, August 15, 2023

Introductory Overview Lecture (IOL) in Interpretable Machine Learning at the Joint Statistical Meeting (JSM), August 8, 2023

5th International Conference on Statistics: Theory and Applications (ICSTA 2023), keynote, August 4, 2023

International Conference on AI and the Digital Economy, keynote, June 26, 2023

Government Advances in Statistical Programming (GASP), keynote, June 14, 2023

INFORMS Conference on Quality, Statistics, and Reliability, North Carolina State University, plenary, June 7, 2023

AI for Scientific Data Analysis: Mini-conference, Chalmers University of Technology, Gothenburg, Sweden, June 1, 2023

Forum for Information Retrieval Evaluation (FIRE), plenary, December 13, 2022

NeurIPS Human in the Loop Learning (HiLL) Workshop, invited talk, December 2, 2022

IEEE/WIC/ACM International Conference on Web Intelligence (WI'22), Keynote, November 18, 2022

Explainable AI in Finance Workshop at the 3rd ACM International Conference on AI in Finance (XAI-FIN@ICAIF2022), Invited Talk, November 2, 2022

Information Technology Laboratory (ITL) Science Day, National Institute for Standards and Technology (NIST), Keynote, October 24, 2022

ECCV Workshop on Safe Artificial Intelligence for Automated Driving (SAIAD), Invited Speaker, October 24, 2022

ECCV Workshop on Responsible Computer Vision, Panel Member, October 23, 2022

Making Sense of Interpretable Machine Learning Workshop, Keynote, Lorentz Center, Leiden, Netherlands, (virtual), October 17, 2022

INFORMS 17th Workshop on Data Mining & Decision Analytics, and **INFORMS 6th Workshop on Data Science**, Keynote, INFORMS, Indianapolis, October 16, 2022

AI-ML Systems Second International Conference on AI-ML Systems, Keynote, October 14, 2022

ICCBR International Conference on Case-Based Reasoning, Keynote, September 14, 2022

ICML workshop on Machine Learning in Healthcare, Keynote, July 23, 2022

ICML workshop on Responsible Decision-Making in Dynamic Environments, Keynote, July 23, 2022

IEEE World AI IOT Congress, Keynote, June 8, 2022

Interpretability in Artificial Intelligence Workshop, Banff International Research Station, Keynote, May 3, 2022

Discover Experiential AI, The Institute for Experiential AI, Northeastern University, Panelist, April 6, 2022

Walmart AI Summit, Plenary speaker, April 6, 2022

INFORMS Optimization Society Conference, Plenary speaker, Greenville, SC, March 14, 2022

5th Annual Swedish Symposium on Deep Learning / 39th Annual Swedish Symposium on Image Analysis, March 14, 2022

Stu Hunter Research Conference on Statistics and Statistical Engineering, 2022, keynote, Duke University, March 8, 2022

AAAI Workshop on Interactive Machine Learning, keynote, February 28, 2022

AAAI Explainable Agency in Artificial Intelligence Workshop, Keynote, February 28, 2022

Government of Canada Data Conference 2022: Driving Data Value and Insights for All Canadians, Panelist, February 23, 2022

Fidelity AI Day, Keynote, February 16, 2022

The Where and Why of Explainable AI, Science and Technology Expert Partnership Advances in Explainable AI Workshop, MITRE Labs, panelist, January 11, 2022

5th Joint International Conference on Data Science & Management of Data (CODS-COMAD 2022), keynote, India (virtual), January 7, 2022

HICSS-55: SWT on Future of Human Work: Harnessing the Power of Augmented Intelligence and Augmented Cognition, panelist, January 3, 2022

Interpretability, Safety, and Security in AI, The Alan Turing Institute, December 14, 2021

Human-Centered AI (HCAI) Workshop at NeurIPS, Keynote, December 13, 2021

Machine Learning & Supply Chain Management Workshop, TRIPODS-X, December 13, 2021

Forward Summit, Puerto Rico Science, Technology & Research Trust, December 10, 2021

2021 IEEE Symposium Series on Computational Intelligence (SSCI), Deep Learning Track, Keynote, December 5, 2021

Gillmore Symposium: Explainable, Interpretable AI: the Future of Investment Management, Warwick University, November 19, 2021

Conference on Non-traditional Data, Machine Learning and Natural Language Processing in Macroeconomics, Bank of Canada, panelist, November 18, 2021

DARPA DSO Futures meeting, plenary speaker, November 18, 2021

Duke in DC discussion on "The Equitable, the Ethical and the Technical: Artificial Intelligence's Role in the U.S. Criminal Justice System," November 15, 2021

McCombs School of Business' Center for Analytics and Transformative Technologies 2021 Global Analytics Conference, keynote, November 12, 2021

Artificial Intelligence in Consumer Finance: Defining and Insuring Fairness, Federal Reserve Bank of Philadelphia and Federal Reserve Bank of Cleveland, November 9, 2021

AAAI Fall Symposium, keynote, November 5, 2021

Advanced Analytics: New Methods and Applications for Macroeconomic Policy, organized by the Bank of England, the European Central Bank, King's College London and King's Business School, keynote, November 4, 2021

Methodological Approaches for Whole Person Research Workshop, NIH, September 29-30, 2021

CNRS School on Explainability, September 30, 2021

iMIMIC Workshop on Interpretability of Machine Intelligence in Medical Image Computing at MICCAI 2021, September 27 2021

Explainable AI Virtual Workshop, Caltech, September 23, 2021

AI Meets Regulators Symposia, AI for Health at Imperial College, September 21, 2021

ICCAI'21 International Conference on Complex Acute Illness, keynote, September 10, 2021

FICO Mastermind Event, Aug 30 – Sept 19, 2021

ARES / CD-MAKE 2021 Conference, keynote, August 20, 2021

ELLIS workshop on Causethical ML, Invited talk, July 26, 2021

ICML Workshop on Theoretic Foundation, Criticism, and Application Trend of Explainable AI, Invited talk, July 23, 2021

CVPR 2021 Tutorial on Interpretable Machine Learning

Forecasting the future for sustainable development, Centre for Excellence and Transdisciplinary Studies, hosted by OECD, keynote, June 17, 2021

Analytics Summit 2021, University of Cincinnati, June 8, 2021

Responsible Machine Learning, keynote, North Carolina State University, June 4, 2021

ICLR Workshop on Responsible AI, May 7, 2021

Bringing Artificial Intelligence to the Bedside, Purdue University Workshop, April 23, 2021

Trustworthy Automated Decision Making (ETAPS 2021 Workshop), keynote, March 28, 2021

Explainable AI Mini-Summit, Re-Work, February 17, 2021

Panel: Bias in AI: How scientists are trying to fix it?, Intuit, February 15, 2021

Florida Women in Mathematics Day, Florida Atlantic University, keynote, February 13, 2021

IEEE EMBS Forum on Data Science and Engineering in Medical Imaging, Symposium #1: Grand Challenges in Data Science and Engineering in Healthcare: Medical Imaging, February 10, 2021

The WACV 2021 Workshop on Explainable & Interpretable Artificial Intelligence for Biometrics (xAI4Biometrics Workshop 2021), keynote, January 5, 2021

Machine Learning in Science & Engineering (MLSE), Columbia University, December 14, 2020

NeurIPS Workshop on Broader Impact of AI, December 12, 2020

Toronto Machine Learning Summit, November 19, 2020

Advancing Analytics 2020, National Conference, Institute of Analytics Professionals of Australia, November 17, 2020

The Machine Learning Conference (MLconf), November 6, 2020

Explainable AI Planning (XAIP) @ ICAPS, October 21, 2020

Symposium on Artificial Intelligence for Learning Health Systems (SAIL), Presymposium, October 21, 2020

Advances in Interpretable Machine Learning and Artificial Intelligence (AIMLAI), keynote, workshop at CIKM, October 20, 2020

Workshop on Credit Card Lending and Payments, Federal Reserve Bank of Philadelphia, September 17, 2020

National Health Symposium, Johns Hopkins Applied Physics Laboratory, September 14, 2020

Data Science, Statistics & Visualization (DSSV), SAMSI, July 29, 2020

Workshop on New Directions in Optimization, Statistics and Machine Learning, Institute for Advanced Study, April 16, 2020

TutORial, INFORMS, October 21, 2019

Advances in Decision Analysis, INFORMS Decision Analysis Society, (keynote), Bocconi University, Milan, June 19, 2019

Debugging Machine Learning Models: ICLR 2019 workshop, May 6, 2019

Safe Machine Learning: Specification, Robustness and Assurance: ICLR 2019 Workshop, May 6th, 2019

FEAP-AI4Fin 2018 : NeurIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: the Impact of Fairness, Explainability, Accuracy, and Privacy, invited talk, December 7, 2018

NeurIPS 2018 Workshop on Critiquing and Correcting Trends in Machine Learning, (Spotlight talk), December 7, 2018

Nicholas Institute, Duke University, invited commenter on deep learning session, October 5, 2018

3rd International Workshop on Biomedical Informatics with Optimization and Machine Learning (BOOM), invited talk, IJCAI workshop, June 13, 2018

Workshop on Human Interpretability in Machine Learning (WHI), invited talk, ICML workshop, June 17, 2018

CVPR NTIRE 2018 New Trends in Image Restoration and Enhancement workshop and challenges on super-resolution, dehazing, and spectral reconstruction, invited talk, representing our competition entry on superresolution, June 18, 2018

Triangle Machine Learning Day, invited talk, April 3, 2018

4th Annual Morgan Stanley Quantitative Equity Research Conference, invited talk, November 17, 2016

Predictive Applications and APIs (PAPIs), October 2016

Japan-America Frontiers of Engineering Symposium, National Academy of Engineering, invited talk, Beckman Center, Irvine CA, June 16, 2016

Criminal Justice in the Age of Big Data, panelist. Harvard Kennedy School. November 13, 2015

Discovery Science, Banff, Canada. October 5, 2015

Duke University Workshop on Sensing and Analysis of High-Dimensional Data, Duke University, July 27-29, 2015

Operations, Technology and Information Management Research Camp, Johnson School of Management, Cornell University. June 23, 2015

MIT Big Data Workshop – "From Data to Insights". May 21, 2015

MIT Media Lab – DARPA ISAT Workshop on Computational Health. April 20, 2015

Workshop on Big Data and Statistical Machine Learning. (Part of the Thematic Program on Statistical Inference, Learning, and Models for Big Data.) Fields Institute, Toronto, January 29, 2015

Bringing Social Science Back In: The 'Big Data' Revolution and Urban Theory, Radcliffe Institute, Harvard University, December 15-16, 2014

Understanding and Improving Cities: Policy/Research Partnerships in the Digital Age, Invited Speaker/Panelist, District Hall, Boston, December 12, 2014

Weathering the Data Storm: The Promise and Challenges of Data Science, Third Annual Symposium on the Future of Computation in Science and Engineering, Harvard University, Jan 26, 2014.

New England Machine Learning Day, May 1, 2013

Exploration and Exploitation 3, Presentation for Second Place team in Data Mining Contest, ICML Workshop, June 2012

IMA Workshop on User-Centered Modeling, May 2012

IMA Workshop on Machine Learning, March 2012

ICML Workshop on Global Challenges, International Conference on Machine Learning, July 2, 2011

American Institute of Mathematics (AIM) workshop on the Mathematics of Ranking, August 16-20, 2010

DIMACS/CCICADA Workshop on Statistical Issues in Analyzing Information from Diverse Sources, Rutgers University, May 6-7, 2010

New England Statistics Symposium (NESS), Harvard University, April 17, 2010

International Utility Working Group: Workshop on Computer-Aided Lean Management (CALM), Columbia University, April 16, 2008

Conference on Machine and Statistical Learning: Prediction and Discovery, AMS-IMS-SIAM Summer Research Conferences in the Mathematical Sciences, invited by organizers Joe Verducci, Xiaotong Shen, and John Lafferty, Snowbird, Utah, June 25-29th, 2006

FOCM 2005 Foundations of Computational Mathematics, Workshop 4 on Learning Theory, Santander Spain. Invited by organizer Steve Smale, 2005

Machine Learning Summer School, Workshop on the Dynamics of Learning, TTI-C, Chicago, invited by organizer Steve Smale, May 16-26, 2005

Notable Panel Discussions

U.S. Senate AI Insight Forum: Transparency, Explainability, Intellectual Property, & Copyright. Invited by Senator Chuck Schumer, facilitated by Leader Schumer, Senator Rounds, Senator Heinrich, and Senator Young, Kennedy Caucus Room at the U.S. Senate, Washington DC, November 30, 2023.

AI Roundtable with Congresswoman Deborah Ross and Congresswoman Valerie Foushee, SAS Campus, November 27, 2023 and February 20, 2024

Invited Presentations at Universities, Societies, and Research Labs

R&D Data Science & Analytics Annual Meeting, PepsiCo Global Research & Development, April 23, 2024
Operations Seminar, Northwestern University, April 15, 2024
Wharton Statistics Seminar, University of Pennsylvania, April 10, 2024
Poetry and Emerging Technologies: A computer scientist and a poet discuss poetry and creativity. Duke University English Department, April 5, 2024
Princeton Quantitative Social Science Colloquium, Princeton University, December 8, 2023
AI for Good Seminar, Organized by International Telecommunication Union (ITU) and United Nations partners, November 15, 2023
Statistics Student Seminar, University of Michigan, November 9, 2023
RAND Statistics Seminar, November 8, 2023
NPL Artificial Intelligence Interest Group, National Physical Laboratory, United Kingdom, May 18, 2023
WNAR Webinar, Western North American Region of The International Biometric Society, May 12, 2023
Harvard Statistics Department Seminar, February 6, 2023
University of Michigan, AI Seminar, January 24, 2023
Women in AI Speaker and Mentorship Series, Deloitte and the Schwartz Reisman Institute for Technology and Society (SRI) at the University of Toronto, January 18, 2023 and January 24, 2023
Tutorial on Machine Learning, Merck, December 12, 2022
Quantitative Biology Seminar Series, Cancer Research UK Cambridge Institute, December 12, 2022
IAA Summit Keynote, Institute for Assured Autonomy, Johns Hopkins University, November 1, 2022
SILO Seminar, University of Wisconsin, October 26, 2022
ATHENA Seminar, Duke University, October 25, 2022
Data, Environments, and Learners: Theory and Algorithms, UCL Statistical Science, October 21, 2022
National Science Foundation (NSF) CISE Distinguished Lecture Series, October 20, 2022
Computational Biomedicine Grand Rounds, Cedars Sinai Medical Center, October 12, 2022
Imageomics External Speaker Series, Imageomics Institute, Ohio State University, October 3, 2022
OIM department seminar at the Wisconsin School of Business, September 30, 2022
NISS/Merck Meetup: Interpretable/Explainable Machine Learning, September 28, 2022
Colloquium, Department of Statistics and Actuarial Science, University of Iowa, September 22, 2022
Keynote for Annual Meeting, Tel Aviv University Center for AI and Data Science (TAD), June 8, 2022
Institute for Advanced Study, Terng Lecturer, Women and Mathematics Program, May 23-27, 2022
Fields Machine Learning Seminar, Fields Institute for Research in Mathematical Sciences University of Toronto, April 25, 2022
Data Valorization Institute (IVADO) in Montreal, Zooming in on Multidisciplinary AI, April 21, 2022
Michigan Institute for Data Science Seminar Series, April 14, 2022
International Conference on Foundations and Applications of AI, Peking University, April 8, 2022
UCSB Center for Responsible Machine Learning Distinguished Lecture Series, UC Santa Barbara, April 8, 2022
"I Can't Believe It's Not Better" Seminar Series, April 7, 2022
Elon University Analytics Day, March 24, 2022
Cambridge Centre for AI in Medicine Seminar, Cambridge UK, March 16, 2022
Janelia Research Campus Computation & Theory Seminar, Janelia Farms, March 10, 2022
The Alan Turing Institute, Human-Machine Teams Seminar, February 25, 2022

Institute for Experiential AI, Northeastern University, Bay Area, Distinguished Speaker Series February 23, 2022
University of Virginia, AI and Machine Learning Seminar, February 4, 2022
Institute for Assured Autonomy Speaker Series, Johns Hopkins University, February 3, 2022
London School of Economics & Political Science, Data Science Seminar, January 31, 2022
Roche Advanced Analytics Network, Roche/Genentech, January 26, 2022
Image Guided Cancer Therapy Seminar, MD Anderson Cancer Center, January 19, 2022
Perspectives in Mathematical Sciences, Dr. F.C. Kohli Centre of Excellence, Chennai Mathematical Institute, January 19, 2022
Trustworthy ML, Fireside Chat, December 16, 2021
The Myth of "Explainable" AI and why "Interpretable" AI is the Answer, Collective AI Podcast, December 15, 2021
AI for Good, Accelerating the United Nations Sustainable Development Goals, Seminar Series, International Telecommunication Union, December 9, 2021
Montreal Speaker Series in the Ethics of AI, December 9, 2021
Chalmers AI Talks, Chalmers University in Sweden, December 8, 2021
Teaching Youth to Build and Deploy Responsible AI for Justice Workshop, Technovation, December 7, 2021
ECE Distinguished Speaker Series, Rice University, December 6, 2021
Oxford Women in Computer Science Distinguished Speaker Series, November 9, 2021
Smarsh Advanced AI panel on Driving Adoption of AI, November 1, 2021
Responsible Modelling in Uncertain Times — CEST-UCL Seminar series, Keynote, November 3, 2021
RBCDSAI LatentView Colloquium, Robert Bosch Centre for Data Science and AI at IIT-Madras, October 29, 2021
TOM Seminar, Harvard Business School, October 23, 2021
Optimization and Data Science Community Seminar, Exxon, October 7, 2021
Neyman Seminar, Berkeley Statistics Department, September 15, 2021
Trustworthy ML Initiative Seminar, Harvard, July 8, 2021
Fidelity AI Seminar, June 29, 2021
University of Basel, Computational Seminar Series, Computational Biology core program of the Biozentrum, Basel, Switzerland, May 31, 2021
University of Oxford, Oxford Computational Statistics and Machine Learning (OxCSML) Seminar, May 14, 2021
FDIC Center for Financial Research Seminar, May 11, 2021
EPFL Statistics Seminar, May 7 2021
Distinguished Seminars on Explainable AI, connected to ERC project "XAI Science and technology for the eXplanation of AI decision making," April 20, 2021
Frontiers of Big Data, AI, and Analytics, Virtual Seminar Series (centered in Australia), April 14, 2021
Iowa State University, Theoretical and Applied Data Science Seminar, April 8, 2021
National Institute of Cancer, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, Seminar, April 7, 2021
New Jersey Institute of Technology Data Science Seminar, March 10, 2021
Metron/George Mason University AI/ML Seminar Series, March 10, 2021
NeEDS Mathematical Optimization Seminar Series, Network of European Data Scientists, February 8, 2021
Brown University, Division of Applied Mathematics Colloquium, February 4, 2021
Texas A&M Institute of Data Science, TAMIDS Seminar, January 30, 2021
Bias^2 Seminar, Harvard Data Science, January 28, 2021
University of Oxford, Computer Science, OATML group, January 26, 2021
University of California, Santa Cruz, Winter Seminar Series, Department of Statistics, January 25, 2021
Verizon Media Research Day, keynote, December 17, 2020
University of Iowa, Department of Business Analytics Seminar Tippie College of Business, December 11, 2020
Duke Computing Roundtable, December 10, 2020
Energy Data Analytics Symposium: Transforming Energy Systems with Data Science Techniques, Duke University, December 9, 2020
Alberta Machine Intelligence Institute, University of Alberta, December 4, 2020
North Carolina Agricultural and Technical State University (NCAT), Seminar, organized by Student Leadership Council and Faculty of the ACIT Institute and TECHLAV Center, November 20, 2020
University of Pittsburgh, Statistics Department Seminar, October 27, 2020
University of California, Davis, Mathematics of Data and Decision in Davis Seminar, October 20, 2020
University of Toronto, MIE Distinguished Seminar Series, Department of Mechanical and Industrial Engineering, October 2, 2020
University of Colorado, Denver, Biostatistics Seminar, September 16, 2020

Online Causal Inference Seminar, August 25, 2020
Tutorial on Interpretable Machine Learning, Joint Statistical Meetings, August 3, 2020
TRIPODS Seminar Series, TRIPODS DATA-INSPIRE Institute, DIMACS & Rutgers CS/Math/Statistics, July 17, 2020
Melbourne Centre for Data Science Seminar Series, June 25, 2020
Decision Making in an Uncertain World, Seminar Series, INFORMS Stochastic Programming Society, June 12, 2020
Center for Human-Compatible AI (CHAI) Seminar, Berkeley, June 10, 2020
Boeing Research Seminar, June 5, 2020
Research Triangle Institute (TRI) Seminar, April 21, 2020
Carnegie Mellon Artificial Intelligence Seminar, April 7, 2020
GSS/SSS webinar program (offered by the American Statistical Association's Government Statistics Section and Social Science Section, February 13, 2020
Ingrid Daubechies Lecture in Computer Science, Duke University, January 21, 2020
North Carolina State University RED Talk, NCSU Data Science Initiative, November 6, 2019
North Carolina Chapter of the American Statistical Association, Webinar, November 1, 2019
Duke University, Langford Lecture, October 31, 2019
North Carolina State University Statistics Department Seminar, October 4, 2019
North Carolina State University ECE Interdisciplinary Distinguished Seminar Series, September 27, 2019
New York University, PRIISM Seminar, September 25, 2019
University of North Carolina at Chapel Hill, Biostatistics Seminar, August 29, 2019
SAMSI Seminar, August 28, 2019
Federal Judicial Center and Duke Law's Bolch Judicial Institute, Duke University, Workshop on Law and Technology for Judges, May 30, 2019
Yale University, MacMillan-CSAP Workshop on Quantitative Research Methods, April 25, 2019
University of Pennsylvania, Mahoney Institute for Neurosciences Seminar, April 8, 2019
Microsoft Research New England Colloquium Series, December 12, 2018
Boston University, Distinguished Speaker Series, Artificial Intelligence Research Initiative, December 10, 2018
Princeton University, Quantitative Social Science Seminar, November 16, 2018
University of North Carolina Statistics and Operations Research Department Colloquium, October 15, 2018
Research Triangle Analysts, lecture October 16, 2018
NSF Webinar: Statistics at a Crossroads: Challenges and Opportunities in the Data Science Era, October 2, 2018
University of North Carolina Health Care and NC Women in Machine Learning and Data Science MeetUp, June 2, 2018
University of North Carolina, Chapel Hill, Causal Inference Research Group Seminar, April 6, 2018
North Carolina State University, Bioinformatics Seminar, March 29, 2018
University of Maryland, Distinguished Speaker Series, Computer Science Department, December 1, 2017
Temple University, Fox School of Business, Statistics Department Colloquium, October 6, 2017
Duke University Algorithms Seminar, September 28, 2017
University of Toronto Law School, Law and Economics Colloquium, September 26, 2017
Microsoft Research, New York City, Sept 19, 2017
University at Buffalo, CSE, Distinguished Speaker Series, May 11, 2017
Columbia University, IEOR-DRO Seminar, May 9, 2017
Statistics C. V. Starr Lectureship Series, Biostatistics Department, Brown University, April 24 2017
Computational Social Science and Public Policy Colloquium, Harris School of Public Policy, University of Chicago, April 14, 2017
Statistics Department Seminar, University of Chicago, April 11, 2017
Duke Network Analysis Center Seminar, Duke University, February 6, 2017
Applied Mathematics Seminar, Duke University, October 3, 2016
Statistics and Complex Systems Seminar, University of Michigan, March 5, 2016
One Day University, New York, April 29, 2016
Urban Social Processes Workshop and Quantitative Methods Workshop, Harvard University, March 10, 2016
Center for Statistics and Machine Learning Seminar Series, Princeton University, October 20, 2015
Applied Statistics Workshop, Harvard University, October 14, 2015.
Brainstorming Session on the Next Generation of Search Engines, Berkman Center for Internet and Society, Harvard University, September 22, 2015
Machine Learning Seminar, Gatsby Unit, University College London. September 16, 2015
MIT Conversation Series, Accenture. July 10, 2015
ENAR webinar, (Eastern North American Region, International Biometric Society), May 8, 2015

Oracle Labs, Research Seminar, April 29, 2015
NYU-Poly Center for Urban Science and Progress, Research Seminar, March 12, 2015
American Express (NYC), Decision Science Seminar, March 11, 2015
Columbia University, IEOR-DRO Distinguished Seminar Series, March 10, 2015
University of Washington, Statistics Seminar, March 2, 2015
Duke University, Machine Learning Seminar, February 23, 2015
Harvard Business School, Technology and Operations Management Seminar, December 19, 2014
NYU Stern School, Department of Information, Operations & Management Sciences, IOMS Colloquium Series, December 3, 2014
UC Berkeley, Seminar in Computer Science Department, November 12, 2014
MIT Lincoln Labs, Seminar, November 4, 2014.
Brown University, Pattern Theory Group Seminar, October 22, 2014
University of Washington, Data Science Seminar, October 15, 2014
University of Alberta, Operations and Information Systems Seminar, October 3, 2014.
Carnegie Mellon University, ECE Seminar / Machine Learning Special Seminar, September 18, 2014
Harvard University, Computer Science Seminar, September 3, 2014
IBM TJ Watson Research Center, KDD Speaker Day, invited talk, August 28, 2014
MIT CSAIL CAP Meeting, May 30, 2014
MIT Lincoln Labs, Seminar, May 27, 2014
Stanford University, Operations Management Seminar, May 6, 2014
Stanford University, Institute for Research in the Social Sciences, Data Science and Inference Seminar, May 5, 2014
UMass Amherst, Machine Learning and Friends Lunch Seminar, April 29, 2014
MIT International Liaison Program Conference, Plenary Speaker, April 23, 2014
Schlumberger-Doll Research Center Seminar, April 14, 2014
Cornell University Operations Research and Industrial Engineering Seminar, April 8, 2014
Liberty Mutual Research Center Seminar, March 14, 2014
University of Pennsylvania Criminology Seminar, March 7, 2014
MIT Theory of Computation Seminar, March 4, 2014
Harvard Applied Statistics Workshop, October 9, 2013
Harvard/MIT Econometrics Workshop, September 12, 2013
MIT Lincoln Laboratory, Seminar, June 4, 2013
Massachusetts General Hospital (MGH), Quantitative Medicine Seminar, April 29, 2013
UT Austin McCombs School of Business, Research Seminar, April 5, 2013
Laboratory for Information Decision Systems (LIDS) lunchtime seminar, February 26, 2013
North Carolina State University, Statistics Department Seminar, February 28, 2013
Yale Statistics Department Seminar, January 14, 2013
Harvard High Dimensional and Correlated Data Seminar, December 17, 2012
MIT Operations Research Center Seminar, December 11, 2012
Yale Computer Science Department Seminar, December 6, 2012
Columbia University, Statistics Department Seminar, October 22, 2012
Robert H. Smith School of Business, University of Maryland, DO&IT Seminar Series, October 12, 2012
MIT Center for Collective Intelligence, July 17, 2012
Notre Dame Computer Science Department Seminar, November 3, 2011
Rutgers Statistics Department Seminar, October 26, 2011
Wharton Statistics Department Seminar, University of Pennsylvania, September 21, 2011
MIT Energy Initiative External Board Meeting Speaker, October 15, 2010
MIT Energy Initiative Seminar Series, October 12, 2010
Boston University, Probability and Statistics Seminar, October 7, 2010
Microsoft Research New England, Machine Learning Seminar, October 4, 2010
Tufts, Computer Science Seminar, September 30, 2010
ABB (Asea Brown Boveri Ltd) Corporate Research Center - United States, Lunch time Seminar, September 1, 2010.
Harvard Statistics Colloquium, April 5, 2010
MIT Operations Research Center Seminar, February 11, 2010
MIT Imaging Seminar, October 22, 2009
University of Chicago, Statistics Department Seminar, February 23, 2009
Ohio State University, Computer Science Department Seminar, February 19, 2009

Brown University, Applied Mathematics Seminar, February 17, 2009
Indiana University, Computer Science Department Seminar, March 5, 2009
University of Houston, Mathematics Department Seminar, Spring 2009
Polytechnic University, Brooklyn, Computer Science Colloquium, April 30, 2007
Columbia University, Applied Math Seminar, April 3, 2007
New York University, Theory Seminar, Computer Science Department, November 9, 2005
IBM Yorktown Heights, April 5, 2005
Rensselaer Polytechnic Institute, March 8, 2005
CCR (IDA Center for Communications Research), Princeton NJ, March 2, 2005
Institute for Advanced Study, Computer Science/Discrete Math Seminar, Princeton, February 14, 2005
Columbia University CCLS Center, February 4, 2005
Google Labs Inc., Research Seminar, October 29, 2004
New York University, Harmonic Analysis and Signal Processing Seminar, October 20, 2004
SUNY at Buffalo, Math Department Seminar, May 2004
NYU, Workshop on Computational and Biological Learning, January 16, 2004

Teaching and Mentoring

- **CS 671D / STA 671D / ECE 687D Machine Learning, Duke**, graduate and undergraduate machine learning, 2016 (Fall), 2018 (Spring), 2019 (Spring), 2019 (Fall), 2020 (Fall), 2021 (Fall), 2022 (Fall), 2023 (Fall)
- **CS 290 / CS 474, Data Science Competition, Duke**, undergraduate course, 2018 (Spring), 2020 (Spring), 2021 (Spring), 2022 (Spring), 2023 (Spring)
- **ME 555 Applications in Data and Materials Science**, (co-taught) 2021 (Spring)
- **Machine Learning Summer School, Duke**, 2018
- **Microsoft-DAT203x, Data Science and Machine Learning Essentials**, co-taught with Stephen Elston, free online course, edX, 2015. Over 17,500 students registered.
- **Microsoft-DAT203.2x Principals of Machine Learning**, co-taught with Stephen Elston, free online course, edX, 2016. Over 14,500 students registered.
- **Microsoft-DAT203.3x, Applied Machine Learning**, co-taught with Stephen Elston, free online course, edX, 2016. Over 8,000 students registered.
- **15.060 Data Models and Decisions, MIT**, MBA course (core course), Fall 2014, Instructor
- **15.075 Statistical Thinking and Data Analysis, MIT**, undergraduate course, Fall 2009, Fall 2010, Fall 2011, Spring 2013, Instructor.
- **15.097 Prediction: Machine Learning and Statistics, MIT**, graduate course, Spring 2012, Instructor. Course materials available on MIT Open Courseware.
- **15.060 Data Models and Decisions, MIT** MBA course (core course), Fall 2012, Instructor
- **15.064 Probability and Statistics, MIT**, Summer 2010, Summer 2011. masters student course (Leaders for Global Operations Program), Co-Instructor, 2010 and 2011
- **COMS 4771 Machine Learning, Computer Science Department, Columbia University**, Spring 2008, lectures on regression, boosting, logistic regression, and ranking.
- **Math 103 Calculus, Princeton**, Fall 2002, Fall 2001, Instructor
  - My lectures were videotaped and placed online. I was the first instructor at Princeton in the sciences to have their lectures videotaped. Class average was over 10 percentage points higher than the average of the other sections on a shared final exam that was worth 50% of their grade; this class was the top scoring class, and it scored 5 percentage points above the second highest class.

- **Math 199 Math Alive, Princeton**, Fall 2003, Teaching Assistant, responsible for the cryptography section, taught by Dr. Ingrid Daubechies
- **Wavelets Course, Program for Women in Mathematics, Institute for Advanced Study**, Summer 2002, Teaching Assistant, taught by Dr. Ingrid Daubechies
- **Physics Classes, Buffalo Seminary Women's High School**, Substitute Teacher, part-time during winter and spring, 1999, taught physics classes daily to freshmen (conceptual physics) and seniors (physics and advanced physics).

Service to Duke

Chair of Committee for Bhuwan Dhingra's reappointment, 2024
Provost's Forum Planning Committee, 2022-2023
Founding faculty member, AI/Materials (aiM) graduate program at Duke 2020-present, graduate admissions committee, 2021-2022
Duke CS Department Graduate Affairs Committee, 2020-2023
Faculty search committees (CS, ECE, and Biostatistics), 2019-2022, 2024
Tenure committee for Prof. David Page, Duke Biostatistics and Bioinformatics, 2020
CS department strategic planning committee, 2019
Tenure Committee for Prof. Kirsten Wickelgren, Duke Math, 2019
Graduate admissions committee, Duke CS, 2019, 2020, 2021
Working group member for white paper "Current State and Near-Term Priorities for AI-Enabled Diagnostic Support Software in Health Care," Duke-Margolis Center for Health Policy
Lead organizer for the Machine Learning seminar, 2016-present
Lead organizer of Triangle Machine Learning Day, 2018, 2019
Chair of faculty search committee, Duke CS/ECE 2017-2018
Bass connections reviewing, 2016
Grad student admissions reviewing 2015-present
Committee for a Prof. Katherine Heller's reappointment 2017-2018
Committee of Guillermo Sapiro, Vince Conitzer, and I, appointed by Provost Kornbluth to write "Computing For Humanity," 2017
Committee of 7 CS/ECE faculty members led by Carlo Tomasi to write a document similar to the above on AI, 2018
Review committee for reappointment of Valerie Ashby, Dean of Trinity College of Arts and Sciences, Duke University, 2018
CS graduate awards committee, 2018
Reviewer for Data+ proposals, 2016-present
Member of numerous RIP, prelim, PhD, MS thesis, and undergraduate honors thesis committees, 2017-2018
Outreach such as hosting Duke Conversations, giving keynote for FEMMES at Duke, meeting with Visiting Committee, etc., 2018-present

Supervision

Postdocs

Dr. Alina Barnett, 2023-present.
Dr. Aaron Fisher, co-advised with Francesca Dominici, Harvard, 2016-2019.
Dr. Keivan Sadeghzadeh, MIT Sloan, 2016.
Dr. Berk Ustun, Harvard CS, 2017-2020. (Now assistant professor at UCSD)
Dr. Noor-E-Alam, MIT Sloan, 2014-2015. (Now assistant professor at Northeastern University)
Dr. Ramin Moghaddass, MIT Sloan, 2013-2015 (Now tenured professor at University of Miami).
Dr. Şeyda Ertekin, MIT Sloan, 2010-2014. (Now assistant professor at Middle East Technical University).

Graduate Students

PhD student Chloe Zhu, Duke CS PhD student, 2024-present.
PhD student Yiyang Wang, Duke ECE MS student, 2023-2024, and PhD student 2024-present.

PhD student Hayden McTavish, Duke CS PhD student, 2023-present.
PhD student Zachary Boner, Duke CS PhD student, 2023-present.
PhD student Gaurav Parikh, Duke CS PhD student, 2023-present.
PhD student Varun Babbar, Duke CS PhD student, 2023-present.
PhD student Eric Chen, Duke CS PhD student, 2023-present.
MS student Julia Yang, Duke ECE MS student, 2023-2024.
MS student Saksham Jain, Duke ECE MS student, 2021-2022.
PhD student Cristina Molero Del Rio, Visiting PhD student, Universidad de Sevilla, 2022-2023.
PhD student Srikar Katta, Duke CS PhD student, 2022-present.
PhD student Jon Donnelly, Duke CS PhD student, 2022-present.
PhD student Sam Rosen, Duke statistics PhD student, 2022-2023.
PhD student Quinn Lanners, Duke Biostatistics PhD student, 2021-present.
MS student Henry Ma, Duke statistics student, 2022-2023.
MD student Sully Chen, Duke Medical Center MD student, 2022-present.
PhD student Kentaro Hoffman, UNC PhD student, 2019-2022.
MS student Pranay Jain, Duke CS MS student, 2021-2023.
PhD student Rui Zhang, Duke CS PhD student, 2021-present.
PhD student Stephen Hahn, Duke ECE PhD student, 2021-present.
PhD student Jiachang Liu, Duke ECE PhD student, 2021-present.
PhD student Zhicheng (Stark) Guo, Duke PhD ECE student, 2021-present.
MS student Xian (Jesse) Sun, Duke MS ECE student, 2020-2021.
MS student Vaishali Jain, Duke MS CS student, 2020-2021.
MS student Bin Han, Duke MS Statistics student, 2019-2020 (now PhD student at University of Washington)
MS student Neha Gupta, Duke MS Economics Computation student and then Duke Economics PhD student, 2019-2024.
PhD student Yingfan Wang (advised as UG, and then Duke PhD student), 2019-present
PhD Student Vittorio Orlandi, Duke Stats student, 2019-2023.
PhD Student Haiyang Huang, Duke CS student, 2019-present.
MS student Henry Yuren Zhang, Duke MS Statistics student 2019-2020.
MS/PhD student Chudi Zhong, Duke MS Statistics student and then Duke CS PhD student, 2019-2023 (now faculty at UNC Chapel Hill)
MS student Jiali Xing, Duke Economics and CS student, 2019-2020.
MS student Matias Benitez Sr., Duke Economics and CS student, 2018-2019.
MS student Chunxiao Li, Duke MS Statistics student, 2018-2019.
MS student Weiyu Yan, Duke ECE student, 2018-2019.
PhD student Usaid Awan, Duke Economics PhD student, 2018-2020.
PhD student Zhi Chen, Duke CS PhD student, 2018-2023.
PhD student Jiayun Dong, Duke Economics PhD student, 2018-2019.
MS student Kangcheng Lin, Duke MS Statistics student, 2018-2019 (now at UIUC PhD program)
MS student Yang Bao, Duke Statistics student, 2018-2019.
MS student Sijia Wang, Duke ECE student, 2018-2019.
MS student Lei Chen, Duke ECE student, 2018-2020.
MS student Xiyang Hu, Duke Statistics student, 2018-2019 (now at CMU PhD program)
PhD Student Alina Barnett, Duke CS student, 2017-2023. (postdoc 2023-2025)
PhD Student Harsh Parikh, Duke CS student, 2018-2023.
MS Student Yameng Liu, Duke Computer Science student, 2017-2019.
PhD Student Lesia Semenova, Duke CS student, 2016-2024 (now at MSR NYC, and then faculty at Rutgers University)
PhD Student Chaofan Chen, Duke CS student, 2016-2020 (now faculty at University of Maine)
PhD Student Tianyu Wang, Duke CS student, 2016-2021 (now faculty at Fudan University)

MS Student Beau Coker, Duke Statistics student, 2017-2018 (now at Harvard PhD program)
PhD Student Marco Morucci, Duke Political Science student, 2017-2021 (now postdoc at NYU)
PhD Student Hongyu Yang, MIT EECS student, 2014-2019.
MS Student Peter Alexander Lee, MIT ORC student, 2015-2016.
MS Student Prashan Wanigasekara, MIT EECS student, 2014-2016.
MS Student Christopher Choo, Engineering and Management, 2014-2015 (now at SUTD and Singapore Grand Prix)
PhD Student Vikas Garg, MIT EECS student, 2014-2016 (co-advised with Tommi Jaakola)
PhD Student Fulton Wang, MIT EECS student, 2013-2018. (Now at Sandia National Labs)
PhD Student Berk Ustun, MIT EECS student, 2012-2017 (Now faculty at UCSD)
PhD Student Stefano Tracà, MIT ORC student, 2012-2018 (now working at Disney Research)
PhD Student Siong Thye Goh, MIT ORC student, 2012-2018.
PhD Student Tong Wang, MIT EECS student, 2012-2016. (Now faculty at University of Iowa)
Project Student Ashia Wilson, MIT Sloan, 2012. (5 months before starting a PhD program at Berkeley)
PhD Student Theja Tulabandhula, MIT EECS student, 2010-2014. (Now senior lecturer at University of Sydney Business School)
PhD Student Ben Letham, MIT ORC student, 2010-2015. (Now at Facebook)
PhD Student Allison Chang, MIT ORC student, co-supervised with Dimitris Bertsimas, 2009-2012 (now at MIT Lincoln Labs).
Masters Student William Harris, MIT ORC Student, co-advised with Michael Ricard, 2014-2015 (now in the US military)
MS Student Oscar Moll, MIT CSAIL student, 2010-2011.
MS Graduate Research Assistant, Nandini Bhardwaj, Columbia & Con Edison Secondary Events Project, 2008.
Masters Project Course, Jawwad Sultan, Columbia & Con Edison Secondary Events Project, Fall 2007.
Summer Students, Supervision of 2 masters students and 1 undergraduate. Columbia & Con Edison Secondary Events Project, Summer 2007.

Undergraduate Students

Duke undergraduate, Chloe Zhu, 2023-2024
Duke undergraduate, Muhang (Tony) Tian, 2023-2024
Duke undergraduate, Danny Luo, 2022-2023.
Duke undergraduate, Boxuan Li, 2022-2023.
Duke undergraduate, Harry Chen, 2022-2023.
Duke undergraduate, Flora Shi, 2022-2023.
Duke undergraduate, Jenny Huang, 2022.
Duke undergraduate, Gaurav Parikh, 2022-2023.
Duke undergraduate, Michelle Qiu, 2022-present.
Duke undergraduate, Anika Mitra, 2022-2023.
Duke undergraduate, Jessie (Yanchen) Ou, 2021-2022.
Duke undergraduate, Jerry Fang, 2021.
Duke undergraduate, Vaibhav Sharma, 2021-2022.
Duke undergraduate, Rui Xin, 2021-2023.
Duke undergraduate, Harsha Srijay, 2021.
Duke undergraduate, Vijit Singh, 2021.
Duke undergraduate, Caleb Kornfeld, 2021-2022.
Duke undergraduate, George Wang, 2021.
Duke undergraduate, Alexander Oesterling, 2021-2022.
Duke undergraduate, Haoning Jiang, 2021-2022.
Duke undergraduate, Lily Zhu, 2021-2023.
Duke undergraduate, Yunyao Zhu, 2021.

Duke undergraduate, Jerry Liu, 2020-2021.
Duke undergraduate, Nathan O'Hara, 2020.
Duke undergraduate, Krystal Hu, 2020.
Duke undergraduate, Angikar Ghosal, 2020-2021.
Duke undergraduate, Thomas Howell, 2020-2021.
Duke undergraduate, Edwin Agnew, 2020-2021.
Duke undergraduate, Benjamin Burnette, 2020-2021.
Duke undergraduate, Jordan Diamond, 2020-2021.
Duke undergraduate, Reed Chen, 2020.
Duke undergraduate, Kari Larson, 2020-2021.
Duke undergraduate, Brandon Zhao, 2019-2021.
Duke undergraduate, Alexander Rubin, 2019.
Duke undergraduate, Feroze Mohideen, 2019.
Duke undergraduate, Diane Hu, 2019-2020.
Duke undergraduate, Isaac Zhang, 2019-2021.
Duke undergraduate, Andre Wang, 2019-2021.
Duke undergraduate, Bhrij Patel, 2019-2021.
Duke undergraduate, Jake Shulman, 2019.
Duke undergraduate, Kenny Green, 2019.
Duke undergraduate, Jerry Pan, 2018-2019.
Duke undergraduate, Alexandru Damian, 2018-2020.
Duke undergraduate, Nikhil Ravi, 2018-2020.
Duke undergraduate, Sachit Menon, 2018-2020.
Duke undergraduate, Chris Suh, 2018-2019.
Duke undergraduate, McCourt Hu, 2018-2019.
Duke undergraduate, Webster Bei, 2018-2020.
Duke undergraduate, Jerry Chia Rui Chang, 2018.
Duke undergraduate, Wilson Zhang, 2018-2019.
Duke undergraduate, Divya Koyyalagunta, 2018-2019.
Duke undergraduate, Anna Sun, 2018-present.
Duke undergraduate, Peter Hase, 2018-2019.
Duke undergraduate, Daniel Tau, 2017-2019.
Duke undergraduate, Caroline Wang, 2017-2020.
Duke undergraduate, Jerry Chia-Rui Chang, 2018.
Duke undergraduate, Hao Liu, summer 2017.
Duke undergraduate, Oscar Li, 2017-2019.
MIT undergraduate, Chelsea Ge, summer 2014.
MIT undergraduate, Jeffrey Chan, spring-fall 2014.
MIT undergraduate, Jiaming Zeng, 2014-2015.
MIT undergraduate, Shawn Qian, summer-fall 2012.
Undergraduate exchange student, Yining Wang, spring 2013.
PhD student at MIT and previously undergraduate from Arizona State University, Lydia Letham, summer 2012, summer 2014.
MIT undergraduate project courses, three students (Kang Zhang, Arash Delijani, Kevin Pang) 2011-2012.
Undergraduate Visiting Student from Ecole Centrale Paris (through MISTI), Fabrice Vegetti, 2012.
Undergraduate Visiting Student from Ecole Centrale Paris (through MISTI), Adel Basli, 2011.
Undergraduate project course on Collaborative Filtering, Association Rules and Information Retrieval, Eugene Kogan, Columbia University, co-supervision with Dr. Ansaf Salleb-Aouissi, Spring 2008.

Undergraduate thesis advisement at Princeton, Krysta Svore, entitled "Multiscale Image Processing Using Single and Double Gaussian Techniques, and Hidden Markov Models," 2001-2002.

Thesis/Prelim Committees (not including my students)

PhD thesis committee for CalTech student, Mechanical and Civil Engineering, 2024
PhD thesis proposal committee and PhD thesis committee for Emory/Georgia Tech student, Cheng Ding, 2023 and 2024
Prelim committee for Duke ECE PhD student, Fakrul Islam Tushar, 2023
Qual committee for Duke MEMS PhD student, Jake Peloquin, 2022
Qual committee for Duke ECE PhD student, Lin Duan, 2021
Prelim and PhD defense committee for Duke CS student Qinwen Huang, 2021, 2023
PhD defense committee for Duke ECE student Haibei Zhu, 2021
Prelim and PhD defense committees for Duke ECE student Mojtaba Zarei, 2021, 2023
Undergraduate honors thesis committee for Yunyao Zhu, 2021
Prelim committee for Duke MEMS student Peiyi Chen, 2021
Prelim and PhD Thesis committee for Duke Biomedical Engineering and Radiology PhD student Yinhao Ren, 2021, 2023
Prelim & PhD thesis committee for Duke ECE PhD student Bohao Huang, 2019, 2020
Prelim committee for Duke MEMS PhD student Bingyin Hu, 2020
Prelim committee for Duke ECE PhD student Jiachang Liu, 2019
Qual committee for Duke ECE PhD student Kavinayan Sivakumar, 2019
Prelim committee for Duke Statistics PhD student Filipe Ossa, 2019
Prelim & PhD Thesis committee for Duke Economics Student Usaid Awan, 2019, 2022
RIP committee for Duke ECE PhD student Jerry Wang, 2019
Prelim committee for Duke Fuqua PhD student Shuyu Chen, 2019
RIP & Prelim committee for Duke CS PhD student Shuai Yuan, 2019, 2020
Qual, Prelim, & PhD committee for Duke ECE PhD student Yuting Ng, 2019, 2021, 2023
Qual committee for Duke ECE PhD student Qian Huang, 2019
Qual, Prelim, & PhD committees for Duke ECE PhD student Ghassen Jerfel, 2019, 2020, 2021
Prelim & PhD thesis committee for Duke ECE PhD student Wanyi Fu, 2019, 2021
Qualifying committee for Duke ECE PhD student Claire Lin, 2019
Prelim committee for Duke CS PhD student Shuzhi Yu, 2018
Prelim and PhD thesis committee for Duke CS PhD student Swarna Ravindran, 2018, 2023
Prelim committee for PhD student Paidamoyo Chapfuwa, ECE PhD student, 2018
RIP committee and prelim committee for Duke CS PhD student Andrew Lee, 2017, 2018
Graduation with honors committee for Duke undergraduate Peter Hase, 2018
Graduation with honors committee for Duke undergraduate Wuming Zhang, 2018
Graduation with honors committee for Duke undergraduate Tianlin Duan, 2018
RIP for Duke PhD student Xiaonan Hu, 2017, 2018
Prelim committee for Duke PhD student Greg Spell, 2018
RIP, Prelim and Thesis committees for Duke PhD student Rachel Draelos, 2017, 2019, 2021
Prelim and thesis committees for Duke PhD student Stavros Sintos, 2017, 2020
RIP and Prelim committee for Duke CS PhD student Zilong Tan, PhD in fall 2018
Thesis committee for Stanford CS student Himabindu Lakkaraju, PhD in spring 2018
RIP committee for Duke CS PhD student Xiaonan Hu, 2017
RIP committee for Duke CS PhD student Shuzhi Yu, 2017
RIP, prelim, PhD committees for Duke CS PhD student Abe Frandsen, 2018, 2019, 2022
Thesis committee for Duke CS undergraduate with distinction Aditya Mukund, 2017
Thesis committee for Duke CS masters student Guan-Wun Hao, 2017
Thesis committee for Duke CS masters student Mona Prakash, 2016
Thesis committee for Duke statistics masters student Emily Shao, 2017
Thesis committee for Duke statistics masters student Sanjay Harihanan, 2017
Thesis committee for Duke ECE Phd student Jordan Hashemi, 2017
Thesis committee for Duke ECE Phd student Zhuoqing Chang, 2017, 2018, 2020
Thesis reader for Harvard CS undergraduate Nicholas Larus-Stone, 2017
Thesis committee for Duke PhD student Shan Shan, 2017

Thesis committee for Duke PhD student Narayanan Rengaswamy, 2017.
Thesis committee for MIT PhD student Yingxiang Yang, 2015.
Thesis committee for MIT PhD student Been Kim, PhD in spring 2015.
Thesis committee for MIT PhD student Anima Singh, PhD in spring 2015.
Thesis committee for Pannaga Shivaswamy at Columbia University CS, PhD in spring 2009.

Society Memberships

-INFORMS
-International Machine Learning Society
-American Statistical Association (ASA)
-Institute of Mathematical Statistics (IMS)
-Association for the Advancement of Artificial Intelligence (AAAI)
-Association for Computing Machinery (ACM)
-American Association for the Advancement of Science (AAAS)
-Society for Causal Inference (SCI)